THREE ESSAYS ON APPLIED ECONOMICS: RURAL ELECTRIC COOPERATIVE CALL CENTER DEMAND, FERTILIZER PRICE RISK, AND ESTIMATING EFFICIENCY WITH DATA AGGREGATION

By

TAEYOON KIM

Bachelor of Science
Seoul National University
Seoul, Korea
2001

Master of Science
Seoul National University
Seoul, Korea
2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2009

THREE ESSAYS ON APPLIED ECONOMICS: RURAL
ELECTRIC COOPERATIVE CALL CENTER DEMAND,
FERTILIZER PRICE RISK, AND ESTIMATING
EFFICIENCY WITH DATA AGGREGATION

Dissertation Approved:

Dr.Philip Kenkel

Dissertation Advisor

Dr. B. Wade Brorsen

Dr. Rodney Holcomb

Dr. Kim Anderson

Dr. Timonthy J. Bowser

Dr. A. Gordon Emslie

Dean of the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

FORORECASTING DEMAND FOR RURAL ELECTRIC COOPERATIVE CALL CENTER

Introduction

The KAMO power (KAMO), an Oklahoma based rural electric cooperative (REC) established an after-hours call center operation for 7 member cooperatives in 2006. The center saved some member cooperatives over $150,000 annually. By 2008 the center had expanded to 18 RECs and KAMO was considering expanding into more states. KAMO contacted Oklahoma State University for assistance in investigating the feasibility of the call center expansion.

The call centers' goal is to answer almost all calls and so it has staff to handle peak calls. Also, call volume is affected by number of customers, season, geographic location and individual REC characteristics such as line maintenance. One of the major challenges is to accurately forecast peak call volume which determines staffing and equipment needs for the call centers' fee structure. In forecasting peak call volume, which is mainly related to the events of ice storms and other disasters, extreme value theory can be applied (Haan and Ferreira, 2006). Modeling total call volume of 18 RECs also raises issues of correlation and dependency. Copula function, which is a dependence function that saparates the marginal distribution from dependence itself, is one approach to modeling the correlations in the call volumes (Cherubini, Luciano, and

Vecchiato, 2004). A centralized call center is expected to generate efficiencies in managing peak call volume because severe weather events which generate the majority of power outage calls do not occur simultaneously across a wide geographic region. Adding additional RECs into KAMO's call center could conceivably reduce peak call volume per member.

The objective of this research is to model call volume for a centralized REC call center and forecast the impact of adding the additional RECs using copula functions. Because the number of calls is large and the underlying disaggregate data are correlated, this study tests several continuous distributions rather than discrete distributions such as a Poission distribution, negative binomial, hurdle Poisson and zero-inflated Poisson which are typically used in modeling call center data (Liu and Cela, 2008). Also, the study illustrates the use of canonical maximum likelihood for predicting extreme values. This technique, which also comes from extreme value theory is appropriate for modeling hourly call volume because the hourly call data have high skewness and dispersion. Because hourly call volume data is not available for the entire study period, peak hourly call volume is estimated conditional on monthly data. Finally, the study compares costs of centralized call center with that of each call center, and examines the effects of adding additional cooperatives with their cost structures.

Call Center's Objective

The objective for operating a call center can be defined as

$$\begin{aligned} &\min_{s_i} \ TC(s_i) \\ \text{(I-1)} \qquad &s.t.\ \text{Prob}[MN(s_i) \geq N_i] \geq \pi \quad \text{for } i = 1, \ldots, I, \\ &\qquad 0 \leq T_{ij} \leq \overline{T} \qquad\qquad\quad \text{for } j = 1, \ldots, J_i, \end{aligned}$$

where TC is the total cost for operating call center, s_i is a staffing level for i^{th} time period, $MN(s_i)$ is the maximum number of calls that are handled by s_i, and N_i is the number of calls that are incoming in time period i, π is a threshold level of satisfying customer needs, e.g., hourly 99th percentile, T_{ij} is a waiting time for j^{th} customer, and $\overline{T}$ is the maximum waiting time, e.g., 10 seconds.

Staffing costs which typically depends on the number of calls anticipated to be received from customers is the largest component of total costs for call center. Forecasting peak and average call volume accurately is essential in order to minimize total costs. A number of approaches to modeling call center volume and staffing needs have been used in previous studies. These include SIPP (stationary, independent, period by period) approach from particular queueing models by Green et al. (2001), and analycal center cutting-plane methods by Atlason et al. (2008). In this study, the staffing requirement for a given number of anticipated calls is based on information from KAMO and historical call volume and staffing data. The maximum number of hourly calls which is handled by a person is assumed to be 110 calls. This is consistent with KAMO's standard operating procedures and is consistent with past results during peak call periods. Also, the focus in this paper is on forecasting peak call volume. Forecasting the peak volume from multiple RECs involves modeling multivariate joint distributions while considering correlation in the pattern of incoming calls from the various cooperatives. The three different regions (Eastern Oklahoma, Western Oklahoma and Missouri) have different topology which could create a different pattern of calls. However, severe weather events such as major ice storms could impact all of the regions leading to correlation of customer calls.

Within agricultural economics, past research has proposed alternative ways to simulated correlated nonnormal random variables. Taylor (1990) provided two procedures for empirically estimating correlated nonnormal joint probability density functions (p.d.f.), and Richardson et al. (2000) also showed how to simulate a multivariate empirical distribution using correlated error terms. These past efforts are heuristic while copulas are based on Sklar's theroem, which says that any joint distribution can be expressed in the form of copula functions.

Copula method

A copula function allows multivariate joint distibutions to have a specific dependency structure. A two-dimensional copula $C(u,v)$ is defined as

(I-2) $C: [0, 1]^2 \rightarrow [0, 1]$.

A copula has the following properties:

$$C(u,0) = C(0,v) = 0 \text{ for all } u,v \in [0, 1],$$

(I-3) $C(u,1) = u$ and $C(1,v) = v$ for all $u,v \in [0, 1]$, and

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0 \text{ for all } u_1 \leq u_2,\ v_1 \leq v_2.$$

The first equation in (I-3) implies that if one of the random variables is 0 then the copula is 0. The second equation implies that if one of random variables is 1 then the copula has the same value as the other variable. The last equation indicates that copula is a non-decreasing function.

According to Sklar's Theorem, any joint distribution $H(x, y)$ with cumulative density functions (c.d.f.) of $F(x)$ and $G(y)$ can be expressed as

(I-4) $H(x, y) = C(F(x), G(y))$,

where $C(\cdot,\cdot)$ is a uniquely determined copula function.

Equation (4) shows that the copula function represents a bivariate joint distribution. If the distribution functions ($F(x), G(y)$) and the copula ($C(\cdot,\cdot)$) are continuous, then equation (I-4) can be restated in terms of the p.d.f. as

(I-5) $h(x, y) = c(F(x), G(y)) \cdot f(x) \cdot g(y)$,

where $h(x, y) = \partial^2 H(x, y) / \partial x \partial y$, $f(x) = \partial F(x)/\partial x$, $g(x) = \partial G(y)/\partial y$, and $c(F(x), G(y)) = \partial^2 C(F(x), G(y)) / \partial F(x) \partial G(y)$ is the copula's density.

Equation (I-5) shows that a marginal distribution ($h(x, y)$) consists of a second-differentiated copula ($c(\cdot,\cdot)$) which has parameters to indicate the dependency structure between two variables (*x, y*) and two probability density functions. Then, canonical representation (Cherubini, Luciano, and Vecchiato, 2004, p.154) for the *n* dimensions can be expressed as

(I-6) $h(x_1, x_2, \ldots, x_n) = c(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \cdot \prod_{i=1}^{n} f_i(x_i)$,

where $c(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) = \partial^n C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) / \partial F_1(x_1) \partial F_2(x_2) \ldots \partial F_n(x_n)$.

Let $\boldsymbol{X} = \{x_{1t}, x_{2t}, \ldots, x_{nt}\}_{t=1}^{T}$ be the sample data matrix, and then the log-likelihood function can be expressed as

(I-7) $l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c(F_1(x_{1t}), F_2(x_{2t}), \ldots, F_n(x_{nt})) + \sum_{t=1}^{T} \sum_{i=1}^{n} \ln f_i(x_{it})$,

where $\boldsymbol{\theta}$ is the set of all parameters of both the marginals and the copula.

Thus, maximum likelihood estimation (MLE) is possible using equation (I-7). Let correlation matrix, $\boldsymbol{R}$, be a symmetric and positive definite matrix with diag($\boldsymbol{R}$) = (1, 1, ... , 1)′ then Gaussian copula can be defined as

(I-8) $c(u_1, u_2, \ldots, u_n) = \frac{1}{|\boldsymbol{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \varsigma'(\boldsymbol{R}_{\text{-1}} - \boldsymbol{I})\varsigma \right)$, where $u_n = F_n(x_{nt})$, $\boldsymbol{R}$ is $n\times n$ correlation matrix, $\varsigma = \left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right)'$, and Φ is the standard normal c.d.f. This is a symetric copula and makes it possible to estimate more than one copula parameter compared to other copulas, e.g., Gumbel copula, Clayton copula, and Frank copula, which have only one parameter and so are a poor choice except in the bivariate case.

Graphical explanation can help readers understand the basic logic of copula function because it is typically defined by using mathmatical notations only. Thus, let's start with 2 random variables following a Gamma distribution and our purpose is to estimate copula parameter. Figure I-1 illustrates the transformation of gamma random variables into standard normal random variables. Let's assume that random variables x_1 and x_2 follow a Gamma distibution. Then, u_1 and u_2 are values for the gamma c.d.f. of x_1 and x_2, respectively, and both range from 0 to 1. Also, $\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$ are inverse of standard normal c.d.f. of u_1 and u_2, respectively, and both are the transformed data of gamma random variables (x_1 and x_2) into standard normal random variables ($\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$) using a Gaussian copula. The correlation between $\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$ is a copula parameter.

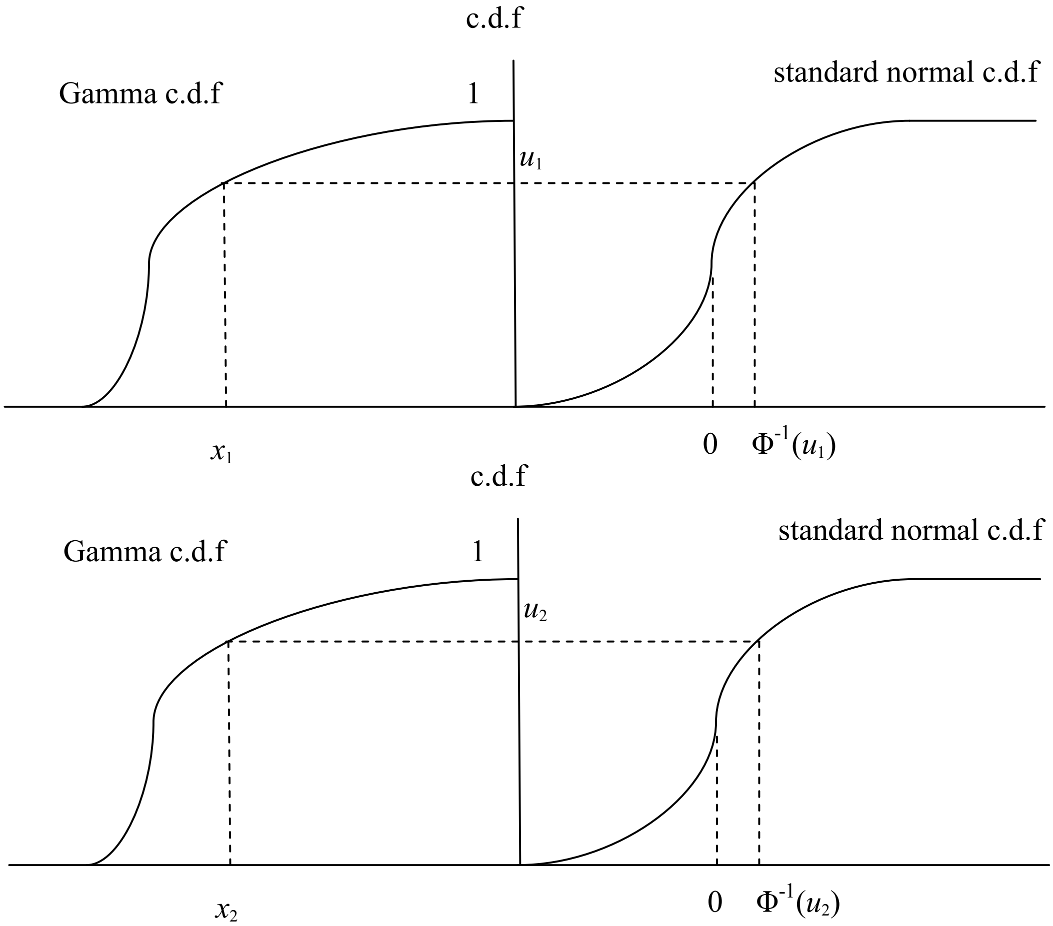


Figure I-1. Graphical depiction of the transformation of gamma random variables into standard normal random variables

Regarding MLE in equation (I-7), it could be computationally difficult to estimate jointly the parameters of the marginal distributions and the parameters of the copula in the case of a high dimension so that Joe and Xu (1996) proposed two steps to estimate $\boldsymbol{\theta}$, which is called the inference for margins (IFM). As a first step, the margins' parameters $\boldsymbol{\theta}_1$ are estimated by usual MLE such as:

(I-9) $$\hat{\boldsymbol{\theta}}_1 = \text{ArgMax}_{\theta_1} \sum_{t=1}^{T} \sum_{i=1}^{n} \ln f_i(x_{it};\boldsymbol{\theta}_1).$$

As a second step, the copula parameters $\boldsymbol{\theta}_2$ are estimated given $\hat{\boldsymbol{\theta}}_1$:

(I-10) $\hat{\boldsymbol{\theta}}_2 = \text{ArgMax}_{\boldsymbol{\theta}_2} \sum_{t=1}^{T} \ln c\left(F_1(x_{1t}), F_2(x_{2t}), \ldots, F_n(x_{nt}); \boldsymbol{\theta}_2, \hat{\boldsymbol{\theta}}_1\right).$

It is known that the IFM estimation is much easier than MLE, particularly when more than 15 parameters are estimated and the IFM estimator, also, is a good starting point for obtaining a MLE estimator (Joe and Xu, 1996).

In terms of fitting marginal distributions, extreme value theory can be applied if modeling the extreme values is the chief concern. Typically, the generalized Pareto distribution (GPD) resulted from the Fisher-Tippett theorem in the generalized extreme value distribution (GEV) is considered (McNeil, 1997) but the empirical distribution is also another possibility (Ghoudi and Rives, 1995). This empirical distribution is used here because other distributions are rejected by several tests such as chi-squared statistic, Kolmogorov-Smirnov statistic, and Anserson-Darling statistic in hourly call volume. The empirical distribution is estimated as the marginal distributions and then copula parameters are estimated by MLE like this:

(I-11) $\hat{\boldsymbol{\theta}}_2 = \text{ArgMax}_{\boldsymbol{\theta}_2} \sum_{t=1}^{T} \ln c\left(\hat{F}_1(x_{1t}), \hat{F}_2(x_{2t}), \ldots, \hat{F}_n(x_{nt}); \boldsymbol{\theta}_2\right),$

where $\hat{F}_i(x_{it})$ is the cumulative distribution calculated from the estimated empirical distribution. This is called the canonical maximum likelihood (CML) (Cherubini, Luciano, and Vecchiato, 2004, p.160).

Simulation can be used to examine the effects of adding additional RECs as well as to predict peak call volume more precisely. Procedures to simulate a Gaussian copula are (Cherubini, Luciano, and Vecchiato, 2004, p.181):

- Find the Cholesky decomposition $\boldsymbol{A}$ of $\boldsymbol{R}$
- Simulate *n* independent random variates $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)'$ from $N(0, 1)$

- Set $\boldsymbol{x} = \boldsymbol{Az}$
- Set $u_i = \Phi(x_i)$ with $i = 1, 2, \ldots, n$, where Φ denotes the univariate standard normal c.d.f.
- $(u_1,\ldots,u_n)' = \left(F_1(y_1),\ldots,F_n(y_n)\right)'$, where $F_i(\cdot)$ is the i^{th} margin
- $F_i^{-1}(u_i) = y_i$, where y_i is i^{th} simulated data from Gaussian copula and its margin

Data and Procedures

Monthly data from January 2006 to June 2008 and hourly data from April 18, 2008 to June 30, 2008 were obtained from KAMO's call center. While hourly data over the entire time period would have been preferred, it was only collected at the end of the time period. Call volume data from 14 RECs which had data for the entire study period were used. In order to capture regional effects which are a main concern for KAMO, call volume data were aggregated over similar regions such as eastern Oklahoma with 7 RECs (Eastern), western Oklahoma with 4 RECs (Western), and south western Missouri with 3 RECs (Missouri). This grouping simplified the use of copula functions because it reduces joint variables into 3 from 14. Since each REC has a different number of meters , call volume per meter rather than total call volume is used. This removes the differences among RECs in terms of their size.

The issue of the distributions being truncated at zero is not present with the monthly call volume data. The minimum montly call volume is 360 calls in the Western region, 766 calls in the Missouri region, and 2,757 calls in the Eastern region. The hourly call volume data does include hours with zero calls creating the issue of truncation at zero. This limited the categories of statistical distributions which could be used to model hourly call volumes.

The @Risk add in program in Microsoft Excel is used to select the best fitting distibution for the marginal distributions of monthly call volume data. The Gamma distibution was shown to have the best fit among several continuous distibutions. Therefore, Gamma distributions and Gaussian copula with IFM method are used for the monthly data.

In terms of marginal distributions for hourly call volume, several distributions like Parato distribution, Gamma distibution, and etc. are tested and they were rejected by several test statistics such as chi-squared statistic, Kolmogorov-Smirnov statistic, and Anserson-Darling statistic. Thus, empirical distributions and Gaussian copula with CML method are selected.

Consider the following Gamma p.d.f.:

$$\text{(I-12)}\quad f(y_t \mid \boldsymbol{x}_t) = \begin{cases} \dfrac{1}{\Gamma(\alpha_t)\beta^{\alpha_t}} y_i^{\alpha_t - 1}\exp(-\dfrac{y_t}{\beta}), & 0 < y_t < \infty \\ 0, & \text{elsewhere,} \end{cases}$$

where $\alpha_t = \underset{\text{t}}{\boldsymbol{x}'}\boldsymbol{\delta}$, $\Gamma(\alpha_t) = \int_0^\infty y_t^{\alpha_t - 1}\exp(-y_t)\,dy_t$, y_t is call volume per month for the t^{th} observation, α_t is the shape parameter which is determined by $\underset{\text{t}}{\boldsymbol{x}'}\boldsymbol{\delta}$, $\boldsymbol{x}_t$ is a vector of explanatory variables, here dummy variables for season, $\boldsymbol{\delta}$ is a vector of unknown parameters to be estimated, and β is the scale parameter to be estimated.

Gaussian copula with 3 random variables can be defined as

$$\text{(I-13)}\quad c(u_1, u_2, u_3) = \frac{1}{|\boldsymbol{R}|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\varsigma'(\boldsymbol{R}_{-1} - \boldsymbol{I})\varsigma\right),$$

where $\boldsymbol{R}$ is 3×3 correlation matrix, $\varsigma = \left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)\right)'$, and Φ is the standard normal c.d.f.

Since there are 3 series of sample data, the log-likelihood function can be expressed as

(I-14) $l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c(u_{1t}, u_{2t}, u_{3t})) + \sum_{t=1}^{T}\sum_{i=1}^{3} \ln f_i(x_{it})$.

Equation (I-14) enables us to use MLE. However, an execution error in SAS procedure is made since it has many parameters such as seasonality parameters of the marginal distributions and copula parameters to estimate. Thus IFM method is used here, i.e., the parameters for gamma distribution such as α, β, and $\boldsymbol{\delta}$ are estimated via MLE using the log-likelihood function from equation (I-12) and then copula parameters in the $\boldsymbol{R}$ matrix given $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\boldsymbol{\delta}}$ are estimated with MLE using the log-likelihood function from equation (I-13).

Consider the following empirical distribution for hourly call volume.

(I-15) $f(y_t) = \begin{cases} \dfrac{1}{T}, & t = 1, \dots, T \\ 0, & \text{elsewhere} \end{cases}$,

where T is the number of observations.

In regard to the CML method for hourly call volume data, the empirical distribution is easily calculated in equation (I-15) and then copula parameters are estimated via MLE from equation (I-16).

(I-16) $\hat{\boldsymbol{\theta}}_2 = \text{ArgMax}_{\boldsymbol{\theta}_2} \sum_{t=1}^{T} \ln c(\hat{u}_{1t}, \hat{u}_{2t}, \hat{u}_{3t}; \boldsymbol{\theta}_2)$.

In order to look at the effect for a certain time period such as 0:00-1:00, 1:00-2:00, … , 23:00-24:00, hourly call volume is separated within each time interval so that coplua parameters and empirical distribution are estimated for each time interval.

Hourly call volume only from April 18 to June 30 in 2008 is another problem when hourly peak (99^{th} percentile) for other month or other season which do not have hourly call volume is regarded. Thus, distribution of hourly data conditional on monthly is considered. Let $H_{[99]}$ be the hourly 99th percentile. Then, using both conditional and marginal distribution, the 99th percentile can be written as

$$\text{(I-17)} \quad \int_0^{H_{[99]}} \int_0^{\infty} f(H|M) f(M)\, dM\, dH = .99\,,$$

where H and M indicate hourly call volume and monthly call volume, respectively.

Since copula correlations between adding additional RECs and existing RECs, and adding additional RECs and other regions do not exist, some assumptions are needed. copula correlations of added RECs with other regions will be assumed as same as the copula correlations of existing RECs with other regions. Also, a single index from average correlations between one of existing RECs and the rest of existing RECs at each region where added RECs come from will be used for copula correlations between added RECs and existing RECs. This single index can be defined as

$$\text{(I-18)} \quad SI_A = \frac{1}{I_A} \sum_{j=1}^{I_A} \frac{\mathbf{Cov}\left(A_j, \left(\left(\sum_{i=1}^{I_A} A_i\right) - A_j \right) \right)}{\mathbf{Var}(A_j)\mathbf{Var}\left(\left(\sum_{i=1}^{I_A} A_i\right) - A_j \right)}\,,$$

where SI_A is a single index for A region, I_A is the total number of RECs for A region, and A_j is the call volume for j^{th} RECs. Average correlations are shown in Tables 5 and 6.

In terms of simulation for call volume, procedures from Cherubini, Luciano, and Vecchiato (2004, p 181) are used to simulate the 99th percentile and effects from adding additional RECs on total number of call volume to the call center. For instance, let's assume that an additional REC with 100,000 members in the Western area is added. Then, a Gamma distribution for the added REC is assumed to have the same parameters as the Gamma distribution for existing RECs in the Western region for monthly call volume. Also, an empirical distirubution for the added REC is assumed to have the same data as the empirical distibution for existing RECs in the Western region for hourly call volume. Copula parameters with Missouri and Eastern regions for added REC are assumed to be the same as those of the existing Western, and copula correlations between added REC and existing Western are assumed to be the same as the average correlation between one of existing RECs and the rest of existing RECs in Western region. Then, hourly 99th percentiles conditional on monthly data of May and June are considered. Here are more specific steps, for example, hourly 99th percentile of time period 10:00-11:00 for summer.

1. Generate M for summer based on the gamma distribution.
2. Generate H for 10:00-11:00 based on the empirical distibution.
3. H conditional on M is calculated using the ratio between M and monthly average from hourly call volume of May and June.
4. Repeat steps 1-3.
5. Calculate 99th percentile of time period 10:00-11:00 for summer with H conditional on M.

Results

Descriptive statistics for monthly and hourly call volume are shown in tables I-1 and I-2, respectively. In hourly call volume, the average of total call is 27 calls per hour. However, the maximum is 612, which indicates high dispersion. Total meters are 262,552 and the Eastern region had the most meters and Missouri had the fewest meters.

Copula correlations for monthly and hourly call volume are shown in tables I-3 and I-4. Copula correlations for monthly call volume are estimated with Gamma distributions and Gaussian copula from equation (I-14) by IFM method while copula correlations for hourly call volume are estimated with empirical distributions and Gaussian copula from equations (I-15) and (I-16) by CML method. It is known that copula correlations capture non-linear associations as well as linear association among correlated data (Cherubini, Luciano, and Vecchiato, 2004, p 38). Here, Pearson correlations which show only linear association among data are reported in monthly call volume and they are slightly different from the copula correlations. Also, most correlations are positive.

Table I-7 indicates hourly 99th percentiles for each time period from hourly data of May and June in 2008, and 10,000 simulated hourly 99th percentiles for summer conditional on monthly data of May and June in 2008 before and after adding an additional REC with 100,000 meters. The highest hourly 99th percentile from hourly data of May and June is 612 while the highest hourly 99th percentile from 10,000 simulated hourly 99th percentiles for summer conditional on monthly data of May and June is 377 for time period 10:00 to 11:00 and much lower than 612. This is because of high monthly call volume of May and June in 2008 compared to the other months shown in Figure I-2. In addition, 377 is hourly 99th percentile for summer not for June, which means that 377 is hourly 99th percentile for June, July and August. As a result, hourly 99th percentile from May and June in 2008 is much higher than hourly 99th percentile for summer conditional on monthly call volume of May and June in 2008. These results show that the high number of calls during the two-month period is a low probability event. If only the hourly data had been used, the staffing recommendations would have been much higher.

Regarding original hourly call volume, hourly 99th percentile, 612, is actual maximum shown in Table I-8 because of 61 observations for a certain time period. From the 10,000 simulation, the maximum conditional on monthly call volume of May and June in 2008 was 757, which is higher than 612. Thus, 612 is actually between 99th percentile and maximum if data has

10,000 observations. From our simulation, 612 was close to 99.9th percentile. The simulated hourly 99th percentile after adding an additional cooperative with 100,000 meters are also shown in Table I-7. The simulated 99th percentiles from adding additional cooperatives are increased and the highest in general for time period 10:00 to 11:00.

Assuming independence among regions is also interesting because it enables us to confirm the usefulness of using copula functions. Same results of 99th percentile compared to Table I-5 are shown in Table I-7. The 99th percentiles assuming independence among regions are lower than those from Table I-5.

Call center personnel are trained to handle calls as expeditiously as possible while being responsive to customer needs. The call center's rule of thumb is that a staff member can handle up to two routine power outage calls per minute. The highest historical call volume in a one hour period was 612 calls which occurred when 6 staff members were present. It therefore appears reasonable to assume that calling-staff members can handle a maximum of 110 calls per hour. Table I-10 indicates that each call center has 1 person for each time period assuming less than 110 calls per hour while staffing for centralized call center varies from 2 to six personnel. Six employees currently staff the KAMO call center from the hours of 17:00 until midnight. Staffing levels are then decreased slightly, reflecting the fact that fewer customers call during the early morning hours. The center is staffed by four employees from midnight until 3:00 and two employees from 3:00 until 8:00 on weekdays. On weekends the center maintains three to six person shifts. The KAMO call center schedule is therefore requiring 2,459 staff person hours per month assuming 21.73 days per month for weekday and 8.69 days per month for weekends while each cooperative's individual call center requires 534 staff person hours per month.

Table I-11 indicates cost comparisons between total of each call center and centralized call center. KAMO call center's average wage for Jannuay 2008 to April 2008 was $13.81/hour. For each call center, the wage is assumed to be $20/hour because of only 1 person to handle after-hours calls. Both are also assumed to have additional benefits of 14.562% . Each telephone line costs $1,200 per year. Additonal costs to create and maintain the radio and software infrastructure are added in the form of depreciation expense including radio, network, and channel bank. Thus annual total cost for 14 call centers is $2,102,104 compared to $486,123 for centralized call center. Cost per meter is $8.01 for 14 call centers compared to $1.85 for centralized call center, which shows huge cost benefits (almost 4 times) of after-hours centralized call center business.

Table I-12 indicates after-hours staffing for centralized call center before and after adding additional cooperatives with 100,000 meters based on 99th percentile. Staffing is determined on the basis of the highest 99th percentile from 0:00 to 3:00, 3:00 to 8:00, 8:00 to 17:00, and 17:00 to 24:00 for each case. Centralized call center requires 1651 staff person hours per month which is lower than actual 2,459 staff person hours per month in Table I-10. In other words, KAMO currenlty copes with almost maximum calls not 99th percentile calls. When added additional cooperatives with 100,000 meters, staffing hours per month from Missouri region, Eastern region, and Western region are 2077, 1973, and 1973, respectively. These are used for cost estimations.

Table I-13 shows cost comparisons between centralized call center before and after adding addtitional cooperatives with 10,000, 50,000, and 100,000 meters from each region. Total costs are calculated in the same way of Table I-11 using staff person hours per month with total salary, telephone line cost, and annual depreciation. In terms of cost per meter, centralized call center has $1.24 based on 99th percentile and it decreases by $0.04 when 10,000 meters are added. The effects from adding 50,000 meters and 100,000 meters are sightly different among regions even though cost per meter decreases compared to cost per meter before adding additional cooperatives. For examples, it decreases by $0.14 from the Missouri region, $0.10 from the Eastern region, and $0.20 from the Western region when 50,000 meters are added. When 100,000 meters are added, it decreases by $0.11 from the Missouri region, $0.16 from the Eastern and Western regions.

Based on 99th percentile, cost per meter, $1.24, before adding additional cooperatives is relatively small compared to actual cost per meter, $1.85, which is based on almost maximum calls per hour. From 10,000 simulation results indicated in Table 8, centralized call center might need to have flexible staffing strategies to cope with 757 call per hour and 967-1008 call per hour with business expansion as maximums.

Conclusions

This research focused on forecasting peak call volume to allow a centralized call center to minimize staffing costs. A Gaussian copula was estimated to capture the correlation among nonnormal distributions. The Gamma distribution was found to provide the best fit for monthly data. The empirical distribution was selected to represent the hourly data based on the ability to represent the 99th percentile in the observed data which had a high degree of positive skewness and dispersion. Hourly 99th percentiles conditional on monthly data of May and June in 2008 were simulated to compensate for lack of hourly data for the entire period.

Estimating peak call volume, simulating data to forecast the 99th percentile, and examining the effects of adding additional cooperative are all important questions for call center managers. These estimates can be easily and more accurately estimated using the marginal probability distribution with the copula function. One result demonstrated by this research is that when positive dependence among data series exists, ignoring their dependence can cause their peak values to be underestimated.

After-hours centralized call center can get cost benefits compared to each call center. The impact of centralized call center for adding an additional cooperative on the peak (99^{th} percentile) call volume affects the cost structure of call center by changing staff person hours per month. The results indicate that after-hours centralized call center costs almost 4 times less than the total of 14 individual call centers. Moreover, when additional cooperatives are added in centralized call center, even though these effects are small, cost per meter decreases up to 16.1% and the degree of cost benefits depends on the regional location of the cooperative.

Centralized call center can currently handle almost maximum calls per hour, which caused cost per meter to be \$1.85. However, it can be reduced into \$1.24 based on 99^{th} percentile. Since maximum call (612) historically came from flooding with severe weather condition in June 1, 2008 in the Eastern region, it can be expected and thus considering flexible staffing strageties would be more beneficial than current staffing.

Generalized Pareto distibution might be better to capture extreme values for monthly call volume than the Gamma distibution. It will be our future research.

Table I-1. Summary Statistics of Monthly Call Volume per Meter from January 2006 to June 2008

Statistics	Total	Missouri	Eastern	Western
Mean	0.04064	0.04040	0.04756	0.02452
	(10,670)	(1,436)	(7,549)	(1,674)
Minimum	0.01503	0.02154	0.01737	0.00527
	(3,945)	(766)	(2,757)	(360)
Maximum	0.09468	0.09280	0.09888	0.08585
	(24,858)	(3,299)	(15,697)	(5,859)
S. D.	0.01626	0.01540	0.01832	0.01637
Skewness	1.44382	1.50094	0.87316	2.29085
Meters	262,552	35,552	158,747	68,253

Note: The numbers of calls are reported in parentheses.

Table I-2. Summary Statistics of Hourly Call Volume per Meter from April 18 to June 30, 2008

Statistics	Total	Missouri	Eastern	Western
Mean	0.000104	0.000095	0.000113	0.000089
	(27)	(3)	(18)	(6)
Minimum	0	0	0	0
	(0)	(0)	(0)	(0)
Maximum	0.002331	0.006385	0.003691	0.002842
	(612)	(227)	(586)	(194)
S. D.	0.00022	0.00034	0.00029	0.00024
Skewness	4.77100	9.99587	6.02059	5.78903
Meters	262,552	35,552	158,747	68,253

Note: The numbers of calls are reported in parentheses.

Table I-3. Copula Correlations with Monthly Call Volume from January 2006 to June 2008

Regions	Missouri	Eastern	Western
Missouri	1 (1)	0.6903 (0.7990)	0.5646 (0.6593)
Eastern	0.6903 (0.7990)	1 (1)	0.7310 (0.7472)
Western	0.5646 (0.6593)	0.7310 (0.7472)	1 (1)

Note: Pearson correlations are reported in parentheses.

Table I-4. Copula Correlations with Hourly Call Volume from May 1 to June 30, 2008

Time Interval	Missouri-Eastern	Missouri-Western	Eastern-Western
0:00-1:00	0.05892	0.23908	0.44568
1:00-2:00	0.30980	0.33467	0.40847
2:00-3:00	0.51006	0.15601	0.37935
3:00-4:00	0.36785	0.16675	0.36110
4:00-5:00	0.13842	0.12214	0.35395
5:00-6:00	0.10203	0.14878	0.27298
6:00-7:00	0.29214	0.12394	0.18252
7:00-8:00	0.35876	0.01733	0.35127
8:00-9:00	0.16884	0.03148	0.66886
9:00-10:00	0.81103	0.66734	0.61745
10:00-11:00	0.69637	0.33174	0.61847
11:00-12:00	0.73122	0.81976	0.67567
12:00-13:00	0.83679	0.76373	0.62991
13:00-14:00	0.51119	0.37387	0.49156
14:00-15:00	0.39940	0.27092	0.47146
15:00-16:00	0.45641	0.51938	0.57839
16:00-17:00	0.11361	-0.29540	0.13798
17:00-18:00	0.40518	0.40307	0.34897
18:00-19:00	0.21247	0.21733	0.19946
19:00-20:00	0.08243	0.05272	0.01376
20:00-21:00	0.12342	0.17851	0.37162
21:00-22:00	0.04309	0.28605	0.23009
22:00-23:00	0.27267	0.41254	0.26393
23:00-24:00	0.51188	0.17701	0.23941

Note: Since copula parameters of diagonal are 1's and copula parameters of off-diagonal are symmetric, Table 4 can be the same format as Table 3 for each time interval.

Table I-5. Houly Average Correlations between One of RECs and the Rest of RECs for Each Region

Regions	Missouri	Eastern	Western
Average Correlations	0.3798	0.5501	0.5708

Table I-6. Monthly Average Correlations between One of RECs and the Rest of RECs for Each Region

Time Interval	Missouri	Eastern	Western
0:00-1:00	0.6976	0.3667	0.0947
1:00-2:00	0.7644	0.2862	-0.0018
2:00-3:00	0.6771	0.3838	-0.0546
3:00-4:00	0.3523	0.5606	0.4525
4:00-5:00	0.5579	0.4644	0.1601
5:00-6:00	0.6379	0.5501	0.0326
6:00-7:00	0.3655	0.5048	0.2184
7:00-8:00	0.2731	0.4039	0.2298
8:00-9:00	0.0507	0.2943	0.1577
9:00-10:00	0.2287	0.3530	0.1933
10:00-11:00	0.6260	0.4739	0.3581
11:00-12:00	0.3570	0.7414	0.3905
12:00-13:00	0.0440	0.6638	0.1757
13:00-14:00	0.0255	0.5067	0.1802
14:00-15:00	0.0039	0.4605	0.1441
15:00-16:00	0.4657	0.3910	-0.0301
16:00-17:00	0.6362	0.2151	0.1276
17:00-18:00	0.4164	0.1827	0.3254
18:00-19:00	-0.0360	0.6437	0.1768
19:00-20:00	0.0052	0.4969	0.4199
20:00-21:00	0.0595	0.4000	0.0134
21:00-22:00	0.2096	0.2992	0.0239
22:00-23:00	0.1850	0.2287	0.1994
23:00-24:00	0.2947	0.1846	0.0439

Table I-7. The Hourly 99th percentile from Hourly Data of May and June in 2008, and the Simulated Hourly 99th Percentile Conditional on Monthly Data of May and June Before and After Adding Additional REC from Each Region with 100,000 Meters for Summer

Time Interval	The hourly 99th percentile from hourly data of May and June	The simulated hourly 99th percentile for summer conditional on monthly data of May and June	The simulated hourly 99th percentile for summer conditional on monthly data of May and June after adding additional rural electric cooperative from each region		
			Missouri	Eastern	Western
0:00-1:00	348	203	468	225	238
1:00-2:00	199	148	348	170	179
2:00-3:00	264	189	240	221	226
3:00-4:00	317	205	235	245	240
4:00-5:00	238	107	157	133	126
5:00-6:00	339	172	197	220	192
6:00-7:00	298	174	235	254	219
7:00-8:00	263	180	219	229	222
8:00-9:00	197	115	127	138	154
9:00-10:00	587	337	353	392	395
10:00-11:00	612	377	409	449	443
11:00-12:00	464	271	284	361	344
12:00-13:00	543	327	377	406	358
13:00-14:00	395	261	283	301	277
14:00-15:00	263	183	356	207	194
15:00-16:00	300	212	269	247	264
16:00-17:00	295	191	228	228	218
17:00-18:00	219	166	193	210	222
18:00-19:00	596	337	366	431	387
19:00-20:00	493	244	259	304	313
20:00-21:00	412	233	269	287	289
21:00-22:00	189	155	170	180	216
22:00-23:00	202	118	145	149	250
23:00-24:00	344	177	330	197	257
Max	612	377	468	449	443

Table I-8. The Hourly 99th percentile and Maximum from Hourly Data of May and June in 2008, and the Simulated Hourly 99th Percentile and Maximum for Summer of Time Period 10:00-11:00 Conditional on Monthly Data of May and June Before and After Adding Additional REC from Each Region with 10000, 50000, and 100000 meters

Time	Hourly Data of May and June	The simulated data conditional on monthly data of May and June	The simulated hourly 99th percentile and maximum conditional on monthly data of May and June after adding additional REC from each reagion w/ 10k, 50k, and 100k meters								
			Missouri			Eastern			Western		
			w/ 10K	w/ 50K	w/ 100K	w/ 10K	w/ 50K	w/ 100K	w/ 10K	w/ 50K	w/ 100K
99th percentile											
May and June	612										
Summer		377	394	402	409	397	421	449	401	420	443
Maximum[1)]											
May and June	612										
Summer		757	967	971	975	967	970	986	971	987	1008

1): 10,000 replications were made.

Table I-9. The Hourly 99th percentile from Hourly Data of May and June in 2008, and the Simulated Hourly 99th Percentile Conditional on Monthly Data of May and June Before and After Adding Additional REC from Each Region with 100,000 meters for Summer Assuming Independence among Regions

Time Interval	The hourly 99th percentile from hourly data of May and June	The simulated hourly 99th percentile for summer conditional on monthly data of May and June	The simulated hourly 99th percentile for summer conditional on monthly data of May and June after adding additional rural electric cooperative from each region		
			Missouri	Eastern	Western
0:00-1:00	348	190	448	202	203
1:00-2:00	199	132	310	138	141
2:00-3:00	264	161	204	185	199
3:00-4:00	317	183	217	203	220
4:00-5:00	238	100	151	119	112
5:00-6:00	339	168	193	184	174
6:00-7:00	298	186	207	206	201
7:00-8:00	263	169	188	194	193
8:00-9:00	197	105	110	118	115
9:00-10:00	587	305	316	350	320
10:00-11:00	612	220	356	400	363
11:00-12:00	464	239	240	274	267
12:00-13:00	543	276	290	319	288
13:00-14:00	395	231	241	266	240
14:00-15:00	263	162	364	175	168
15:00-16:00	300	180	226	200	230
16:00-17:00	295	185	221	217	212
17:00-18:00	219	157	169	180	191
18:00-19:00	596	327	340	381	347
19:00-20:00	493	237	253	275	314
20:00-21:00	412	225	257	262	262
21:00-22:00	189	152	162	162	198
22:00-23:00	202	107	119	122	238
23:00-24:00	344	137	289	150	225
Max	612	327	448	400	363

Table I-10. After-Hours Staffing for Each Call Center and Centralized Call Center

Time	Each Call Center		Centralized Call Center	
	Weekday	Weekends	Weekday	Weekends
0:00-1:00	1	1	4	4
1:00-2:00	1	1	4	4
2:00-3:00	1	1	4	4
3:00-4:00	1	1	2	3
4:00-5:00	1	1	2	3
5:00-6:00	1	1	2	3
6:00-7:00	1	1	2	3
7:00-8:00	1	1	2	3
8:00-9:00		1		6
9:00-10:00		1		6
10:00-11:00		1		6
11:00-12:00		1		6
12:00-13:00		1		6
13:00-14:00		1		6
14:00-15:00		1		6
15:00-16:00		1		6
16:00-17:00		1		6
17:00-18:00	1	1	6	6
18:00-19:00	1	1	6	6
19:00-20:00	1	1	6	6
20:00-21:00	1	1	6	6
21:00-22:00	1	1	6	6
22:00-23:00	1	1	6	6
23:00-24:00	1	1	6	6
Sum	15	24	64	123
Average days per Month	21.73	8.69	21.73	8.69
Staffing hours per Month (15*21.73+24*8.69 vs. 64*21.73+123*8.69)	534		2459	

Table I-11. Cost Comparisons between Total of Each Call Center and Centralized Call Center

Cost	Total of Each Call Center	Centralized Call Center
Salary ($20/hour*534hour[1)]*14 coops*12 months vs. $13.81/hour*2459hour[1)]*1 coop*12 months)	$1,795,800	$407,573
Other benefits (14.562%)	$261,504	$59,351
Total salary (Salary + Other benefits)	$2,057,304	$466,923
Telephone line ($1200*1*14 vs. $1200*6)	$16,800	$7,200
Radio ($3000*1*14 vs. $3000*6)	$42,000	$18,000
Network ($2500*1*14 vs. $2500*6)	$35,000	$15,000
Channel bank ($4500*1*14 vs. $4500*6)	$63,000	$27,000
Annual depreciation (5 year)	$28,000	$12,000
Total cost (Total salary + Telephone line + Annual depreciation)	$2,102,104	$486,123
Total meters	262,552	262,552
Cost per meters	$8.01	$1.85

Note: 1) Both 534 hour and 2459 hour come from Table 10.

Table I-12. After-Hours Staffing for Centralized Call Center before and after Adding Additional RECs with 100,000 Meters Based on 99th Percentile

Time	Before Adding Additional RECs		After Adding Additional RECs					
			Missouri		Eastern		Western	
	weekday	weekends	weekday	weekends	weekday	weekends	weekday	weekends
0:00-1:00	2	2	5	5	3	3	3	3
1:00-2:00	2	2	5	5	3	3	3	3
2:00-3:00	2	2	5	5	3	3	3	3
3:00-4:00	2	2	3	3	3	3	3	3
4:00-5:00	2	2	3	3	3	3	3	3
5:00-6:00	2	2	3	3	3	3	3	3
6:00-7:00	2	2	3	3	3	3	3	3
7:00-8:00	2	2	3	3	3	3	3	3
8:00-9:00		4		4		5		5
9:00-10:00		4		4		5		5
10:00-11:00		4		4		5		5
11:00-12:00		4		4		5		5
12:00-13:00		4		4		5		5
13:00-14:00		4		4		5		5
14:00-15:00		4		4		5		5
15:00-16:00		4		4		5		5
16:00-17:00		4		4		5		5
17:00-18:00	4	4	4	4	4	4	4	4
18:00-19:00	4	4	4	4	4	4	4	4
19:00-20:00	4	4	4	4	4	4	4	4
20:00-21:00	4	4	4	4	4	4	4	4
21:00-22:00	4	4	4	4	4	4	4	4
22:00-23:00	4	4	4	4	4	4	4	4
23:00-24:00	4	4	4	4	4	4	4	4
Sum	44	80	58	94	52	97	52	97
Staffing hours per month	1651		2077		1973		1973	

Table I-13. Cost Comparisons for Centralized Call Center before and after Adding Additional REC from Each Region with 10000, 50000, 100000 meters

Centralized Call Center	Total meters	Total cost	Cost per meter	Cost Savings (%)
Based on 99th percentile	262552	$326,282	$1.24	
Adding additional coops				
from Missouri				
with 10000 meters	272552	$326,282	$1.20	3.2%
with 50000 meters	312552	$343,606	$1.10	11.3%
with 100000 meters	362552	$410,327	$1.13	8.9%
from Eastern				
with 10000 meters	272552	$326,282	$1.20	3.2%
with 50000 meters	312552	$355,155	$1.14	8.1%
with 100000 meters	362552	$390,529	$1.08	12.9%
from Western				
with 10000 meters	272552	$326,282	$1.20	3.2%
with 50000 meters	312552	$326,282	$1.04	16.1%
with 100000 meters	362552	$390,529	$1.08	12.9%

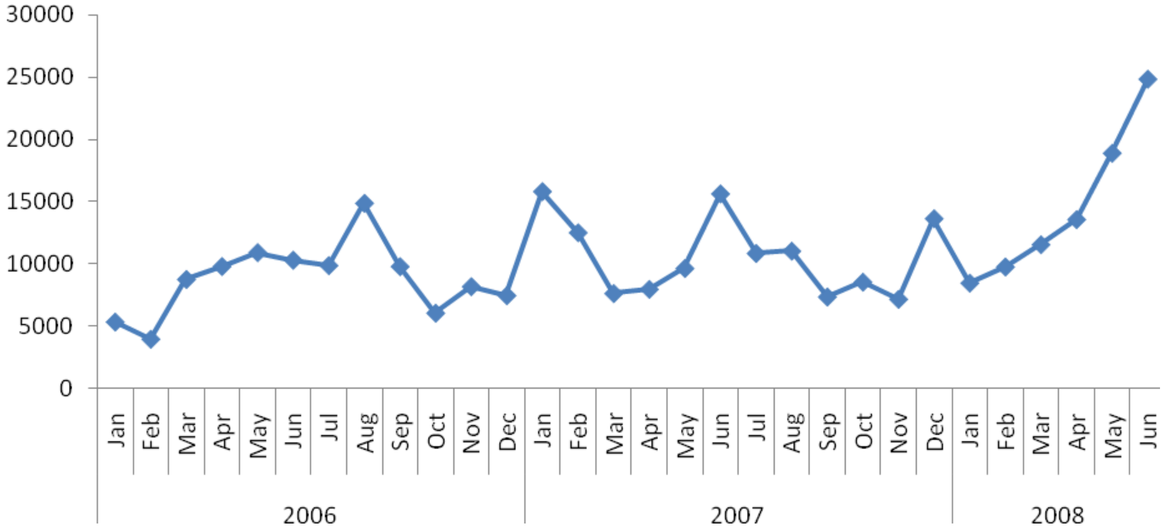


Figure I-2. **Monthly call volume from January 2006 to June 2008**

CHAPTER II

HEDGING EFFECTIVENESS AND OPTIMAL CASH PURCHASE STRATEGIES FOR FERTILIZER IN OKLAHOMA

Introduction

Fertilizer costs account for 30-35% of the variable production costs for dry land Oklahoma wheat production and up to 85% of variable expenses for some forage crops (Oklahoma State University). Between the spring of 2002 and the spring of 2008 fertilizer prices increased dramatically. The farm-level price of nitrogen fertilizer increased 300-375% and increased 95% during 2007-2008 (USDA). The price of phosphate products increased 400% during 2002-2008 and almost 100% during 2007-2008. During the summer of 2008 fertilizer prices decreased dramatically with both nitrogen and phosphate products falling over 50% (Laws). These price changes demonstrate the high volatility in fertilizer prices. Historically, the price variation within a marketing year was $15-$20/ton. Price changes for anhydrous ammonia fertilizer of $100/ton have occurred during the last three seasons and price levels for both nitrogen and phosphate products have changed over $500/ton during the past 12 months.

Fertilizer price volatility has impacted all levels of the supply chain. Historically, farm supply firms have attempted to limit retail price volatility, covering the risk of price swings in their margin structure. Supply firms have also historically stockpiled fertilizer for peak demand periods. Because of increasing price volatility, the timing of fertilizer purchases has become a major risk factor for fertilizer dealers. Some dealers are attempting to shift price risk to producers through advance purchases. Both producers and fertilizer retailers are seeking new strategies to manage fertilizer price risk. Some of these strategies involve decision rules for timing fertilizer purchases.

Risk management options with futures or over the counter (OTC) derivatives are limited. Fertilizer contracts on the Chicago Mercantile Exchange (CME) were discontinued due to a lack of liquidity (Bollman, Garcia and Thompson) while transactions on the Direct Hedge Exchange, based in Switzerland, has a 5,000 ton contract size for fertilizer that is not workable for many retailers, much less producers. OTC strategies require relationships with a brokerage firm or OTC derivative provider and the expertise to manage the required transactions. Basis risk, the difference between the closing futures market contract price and the farm level price for fertilizer, can also be substantial (Bollman, Garcia and Thompson, 1996). Cross hedging fertilizer with natural gas contracts has been found to be ineffective (Dhuyvetter, Albright and Purcell, 2001). Because opportunities to control fertilizer price risk through futures market instruments are limited, dealers and producers rely on cash purchase and storage strategies to manage price risk. Cash purchase strategies attempt to diversify price risk by distributing purchases across the year and/or timing purchases to take advantage of seasonal price movements. Every fertilizer dealer or producer who inventories fertilizer must implement some strategy for purchasing their fertilizer inventory. The main objective of this research is to determine the effect of strategies to systematically purchase fertilizer at pre-determined calendar periods on the average level and year-to-year variability of fertilizer prices. In order to provide a benchmark for comparison, the historical hedging effectiveness and optimal hedging ratio for fertilizers based on cash prices of Oklahoma and futures prices of (the now discontinued) CME fertilizer contracts are estimated.

Theory

Theoretical foundation for hedging starts from Johnson (1960) and Stein (1961), and it is followed by Ederington (1979). Minimum variance hedge ratios can be derived from the following profit equation

(II-1) $\pi_h = X_s(p_1 - p_0) + X_f(f_1 - f_0)$

where π_h is a profit for hedged firm, X_s is spot position, X_f is future position, p_0 and f_0 are initial spot and future prices, and p_1 and f_1 are unknown ending spot and futures prices. If the unknown spot and futures prices are only treated as random variables, then the variance of the profit can be expressed as

(II-2) $\mathrm{Var}(\pi_h) = X_s^2\mathrm{Var}(p_1 - p_0) + X_f^2\mathrm{Var}(f_1 - f_0) + 2X_s X_f\mathrm{Cov}(p_1 - p_0, f_1 - f_0)$.

Solving X_f by minimizing the equation (II-2) gives the optimal futures position for the minimum variance such as

(II-3) $X_f^* = -X_s\ \mathrm{Cov}(p_1 - p_0, f_1 - f_0)/\mathrm{Var}(f_1 - f_0)$.

Then, the minimum variance hedge ratio (X_f^*/X_s) is the regression slope of futures price changes against spot price changes.

When $X_f = 0$, an unhedged profit is $\pi_u = X_s(p_1 - p_0)$ and the variance of unhedged profit is $\mathrm{Var}(\pi_u) = X_s^2\mathrm{Var}(p_1 - p_0)$. Hedging effectiveness (Ederington, 1979) can be defined as the percent reduction in the variance and thus it is identical to the correlation (ρ) between spot and futures price changes such as

(II-4) $e = 1 - \mathbf{Var}(\pi_h)/\mathbf{Var}(\pi_u) = \rho$

where e is a hedging effectiveness.

Data and Procedures

A 17 year time series of weekly fertilizer spot prices from Jan. 7, 1991 to Oct. 29, 2007 at two Oklahoma delivery points (Enid Oklahoma and the Tulsa Port of Catoosa) were obtained from fertilizer industry sources. Enid Oklahoma is in the center of the Oklahoma wheat belt and receives fertilizer by truck and rail. The Tulsa Port of Catoosa is located on the McClellan-Kerr Arkansas River Navigation System which received barge shipments from New Orleans and other ports on the Gulf of Mexico. The data included prices for the three nitrogen products (NH3, urea, and UAN) and one phosphorus formulation (DAP) that constitute the majority of Oklahoma fertilizer products. In order to look at cash market strategies, spot prices within each year were adjusted to reflect the interest costs associated with purchase and storage strategies. The fixed costs of warehouse ownership were not considered.

In order to provide a benchmark for comparing the effectiveness of cash purchase and future market strategies, a 4 year time series of weekly fertilizer future prices from Jun. 7, 2004 to Jul. 9, 2007 (the date the contracts were de-listed) were obtained from the CME. Hedging effectiveness and optimal hedging ratio for urea, UAN and DAP fertilizer forms were determined by linearly regressing price differences in future and spot market data based on the equation (3) and (4). Hedging effectiveness was determined for weekly hedges, reflecting the situation of a farm supply dealer that who places hedges every week of the year, and for monthly hedges which reflected a strategy of placing hedges and making cash purchases 12 times a year.

The cash market purchasing strategies including a number of scenarios of systematically purchasing fertilizer on a consistent calendar date each year. The cash market strategies were analyzed using historical price data. The scenarios considered were designed to represent alternatives available for a typical fertilizer dealer or a producer who has the ability to store fertilizer. Annual fertilizer usage was assumed to be split evenly across fall and spring application seasons. Spring application of fertilizer was assumed to occur in the first week of February while the fall application period was timed for the second week of August. Warehouse capacity constraints of 25%, 50%, 75% and 100% of annual usage were incorporated into the scenarios considered. The remaining amount of fertilizer usage in excess of warehouse capacity was assumed to be purchased during the application season.

Seventeen scenarios were analyzed. The first four scenarios represented the minimum average fertilizer price that could be achieved by systematically purchasing 25%, 50%, 75% and 100% of annual fertilizer usage in advance of the application period at a consistent annual date or dates selected by the model. The second four scenarios were similar except that the consistent annual day or dates were selected to minimize the variance in fertilizer price over the historical data period. In order to provide a benchmark as to the possible impact of purchase dates on price and variance, eight companion scenarios representing the purchase dates generating the maximum average price and maximum variance at each warehouse capacity constraint were also included. Another value of these scenarios is that they identify the time periods that dealers and producers should avoid purchasing fertilizer. The final scenario consisted of purchasing a even amount of annual usage during every week of the year. This scenario provided a benchmark by which the other sixteen scenarios could be measured.

An optimization model from the Excel program with solver option was used to determine the purchase date (week of the year) or multiple dates that minimized or maximized the average spot price or price variance for the 17 year period. The purchase date selected by the model was applied to the entire 17 year price series. The selected dates represented mechanical cash purchase strategies that could be used by a fertilizer dealer or producer.

Results

Hedging effectiveness and minimum hedge ratios for Urea, UAN, and DAP are reported in Tables II-1, 2, and 3. Two hedging periods are considered here such as week hedges and month hedges in order to look at those differences. In general, hedging effectiveness for monthly hedges are greater than those for weekly hedges, which is consistent with Ederington's findings (1979). Monthly hedges are more typically more effective because future prices have more time to respond to changes in supply and demand and other market factors. Hedging with the nearby contract was more effective for Urea while there are no patterns for UAN and DAP.

In terms of locations, the effectiveness of monthly hedges of Urea and UAN was greater for Tulsa than for Enid. As a river market, the nitrogen products sold at the Tulsa port were transported from New Orleans import ports and are thus related to the Urea and UAN contract delivery points. Enid is an inland manufacturing point for Urea and UAN with prices less related to the New Orleans market. Hedging effectiveness for DAP is examined for Tulsa. Almost all DAP sold in Oklahoma is sourced off of the Tulsa (Arkansas river) port market. The DAP delivered to Tulsa comes from international sources delivered to New Orleans ports and from domestic production in Central Florida.

Placing monthly hedges for urea would have reduced price variance by 50% for a farm supply dealer purchasing urea from the Tulsa market and 44% for a dealer who sourced off of the Enid manufacturing point. The effectiveness of monthly hedges for UAN was 28% for the Tulsa market and 21% for the Enid market. The effectiveness of monthly hedges for DAP was 26% for the Tulsa market. Two factors are likely to create the relative ineffectiveness of the fertilizer hedges. First, the Oklahoma markets are removed from and therefore somewhat uncorrelated with the contract delivery points (New Orleans for UAN and urea, and Central Florida for DAP). Second, the fertilizer contracts were thinly traded and thus price convergence was poor even at the delivery points. Tables II-1-3 also show the hedging effectiveness at the New Orleans and Central Florida markets. The effectiveness of monthly hedges relative to the New Orleans cash market was 65% for urea and 30% for UAN. The effectiveness of monthly hedges relative to the Central Florida cash market was 30% for DAP. In light of these results, it is not surprising that the CME fertilizer contracts were discontinued. This has left fertilizer dealers in Oklahoma and other locations with cash purchase strategies as the major remaining tool for risk management.

The impacts of the mechanical purchase strategies selected by the model in reducing the average spot price of fertilizer are shown in Table II-6. Compared to a base strategy of purchasing an even amount of fertilizer each week, systematically purchasing during the weeks selected by the model reduced the average fertilizer price by 3-7%. The results generally showed a benefit from increased warehouse capacity. There was no additional benefit of increasing warehouse capacity from 75% to 100% of annual needs for some product forms. The mechanical purchase strategies had a greater impact on the average price of the UAN formulation relative to Urea or DAP. The results were similar for the Tulsa (river port) and Enid (inland manufacturing point) location.

The impact of the mechanical cash purchase strategies designed to reduce the year-to-year variance in fertilizer prices are shown in Table II-7. While the baseline strategy of purchasing an even amount each week might be expected to reduce the year-to-year price variation, the results indicated further benefits from purchasing on the dates selected by the model. The results were most dramatic for DAP where year-to-year variance assuming 100% warehouse capacity was only 43% of that of the baseline scenario. Not surprisingly, additional warehouse capacity increased the ability to reduce price variance. A dealer with warehouse capacity equal to annual usage could reduce price risk by 57% for DAP, 33% or UAN and 17 to 26% for urea. This indicates that, for the context of an Oklahoma fertilizer dealer, cash purchase strategies are more effective in reducing price variance relative to the historical hedging strategies for UAN and DAP but less effective for urea.

The difference between the average price resulting from purchase dates selected to generate the minimum fertilizer price and the average price resulting from purchase dates selected to generate the maximum fertilizer price are shown in Table II-8. As before, the prices are shown as an index relative to price resulting from purchasing an even amount each week. The results in Table 8 help to answer the question "how important are purchase dates on determining the average cost of fertilizer?" The results indicated that there was a substantial difference (7% to 16%) difference in fertilizer price between a fertilizer dealer that had systematically purchased during the highest price dates relative to a dealer who systematically purchased on the lowest price dates. These results suggest that fertilizer dealers can use historical data to determine the optimal dates to purchase fertilizer and to identify calendar periods during which they should avoid purchases.

The difference in the variance in fertilizer prices between the prices created by systematically purchasing on the dates selected to minimize and maximize variance is provided in Table II-9. The results indicated that the timing of fertilizer purchases has a major impact on the year to year variation in fertilizer price. The difference in variance ranged from 21% to 83% depending on warehouse constraint and product form. The results highlight the importance of purchase dates on the year-to-year variance in fertilizer prices.

The optimal purchase dates identified by the model for the various objectives, product forms and locations are provided in Table II-10. In the case of Urea products purchasing in mid-summer achieved the lowest average price. Purchasing in spring yielded the highest price. For the UAN formulation purchasing in mid-November minimized price while purchasing in late April resulted in the highest price. The seasonal price patterns for DAP were similar with early November being the best time to purchase and late March was, on average, the worst date. The purchase date or dates which minimized the year-to-year variation in fertilizer price are more difficult to characterize. Spreading purchases at numerous systematic dates throughout the year minimized variance for Urea while purchasing in January and November achieved the lowest variance for the other formulations. Purchasing in mid-fall (Urea-Tulsa) or late spring (other formulation and locations) resulted in the greatest year-to-year variation in prices.

Conclusions and Implications

The level and volatility of fertilizer prices is an area of great concern for producers and agribusiness firms. Hedging and option strategies to manage fertilizer price risks are limited because they had low hedging effectiveness and now their futures are not available. As a bench mark, this study estimated the hedging effectiveness and optimal hedging ratio for fertilizer relative to Oklahoma cash markets. The results demonstrated that hedging fertilizer purchases was never a very effective risk management tool for Oklahoma fertilizer dealers, reducing variance by only 44-50% for urea, 21-28% or UAN and 26-30% for DAP.

The focus of this study was on examining the success of mechanical cash purchase strategies in reducing the average price or year-to-year price variability of fertilizer. Mechanical purchase strategies which involved purchasing and inventorying fertilizer in mid-summer or in late fall would have historically reduced the average price by 4 to7% and variance by 17 to 57% depending on product form and cash market location. For Oklahoma fertilizer dealers, cash purchase strategies are more effective in reducing price variance relative to the historical hedging strategies for UAN and DAP but less effective for urea. While the results did not consider the fixed costs of warehouse ownership, increasing warehouse capacity had a significant impact on the effectiveness of strategies to decrease fertilizer price or variance.

In considering these results, two important limitations of the study should be emphasized. First, the fertilizer price data used represented locations in Oklahoma. The application periods were modeled to represent the requirements of winter wheat. Fertilizer price patterns are likely affected by the usage in the corn belt. The optimal purchase times for dealers and producers in Oklahoma appear to be the time periods out of cycle with corn belt usage. Dealers and producers in the mid-west, analyzing historical prices for their locations might find it more difficult to identify effective cash purchase strategies. The second limitation is that the results are based on seventeen years of fertilizer price data. The purchase periods identified by the model represent the time periods that historically would have reduced fertilizer price or variance. The fertilizer supply chain has undergone significant structural change. Seasonal price patterns identified in the historical data may not extend to future periods. Nevertheless, these results provide a logical starting point for a fertilizer dealer or producer determining the timing of fertilizer purchases.

Table II-1. Hedging Effectiveness and Optimal Hedging Ratio for Urea

Hedge Period	The Futures Contract	Hedging Effectiveness	Optimal Hedging Ratio
Week Hedges	Tulsa (153 observations)		
	The Nearby Contract(1-3 Month)	0.2637	0.2830***
	3-6 Month Contract	-0.0850	-0.0789
	6-9 Month Contract	0.0937	0.0612
	Enid (153 observations)		
	The Nearby Contract(1-3 Month)	0.3016	0.4168***
	3-6 Month Contract	-0.0488	-0.0597
	6-9 Month Contract	0.0052	0.0045
	New Orleans and L.A. (153 observations)		
	The Nearby Contract(1-3 Month)	0.3502	0.3629***
	3-6 Month Contract	-0.0245	-0.0240
	6-9 Month Contract	0.0936	0.0638
Month Hedges	Tulsa (37, 35, and 29 observations)		
	The Nearby Contract(1-3 Month)	0.4977	0.6577***
	3-6 Month Contract	0.1559	0.2207
	6-9 Month Contract	0.3602	0.4100*
	Enid (37, 35, and 29 observations)		
	The Nearby Contract(1-3 Month)	0.4350	0.6159***
	3-6 Month Contract	0.0475	0.0730
	6-9 Month Contract	0.3607	0.4451*
	New Orleans and L.A. (37, 35, and 29 observations)		
	The Nearby Contract(1-3 Month)	0.6487	0.8912***
	3-6 Month Contract	0.3772	0.55482**
	6-9 Month Contract	0.5602	0.6769***

Note: Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance at 10%, 5%, and 1%, respectively.

Table II-2. Hedging Effectiveness and Optimal Hedging Ratio for UAN

Hedge Period	The Futures Contract	Hedging Effectiveness	Optimal Hedging Ratio
Week Hedges	Tulsa (130 observations)		
	The Nearby Contract(1-3 Month)	0.0721	0.0889
	3-6 Month Contract	-0.0190	-0.0219
	6-9 Month Contract	-0.0212	-0.0196
	Enid (130 observations)		
	The Nearby Contract(1-3 Month)	0.0476	0.0538
	3-6 Month Contract	0.0203	0.0235
	6-9 Month Contract	0.1004	0.0889
	New Orleans and L.A. (130 observations)		
	The Nearby Contract(1-3 Month)	0.1387	0.1527*
	3-6 Month Contract	0.1787	0.1787**
	6-9 Month Contract	0.2550	0.2368***
Month Hedges	Tulsa (37, 35, and 33 observations)		
	The Nearby Contract(1-3 Month)	0.2764	0.3309*
	3-6 Month Contract	0.1605	0.1743
	6-9 Month Contract	0.1207	0.1049
	Enid (37, 35, and 33 observations)		
	The Nearby Contract(1-3 Month)	0.2095	0.2403
	3-6 Month Contract	0.1828	0.2071
	6-9 Month Contract	0.4461	0.3701***
	New Orleans and L.A. (37, 35, and 33 observations)		
	The Nearby Contract(1-3 Month)	0.2998	0.4046*
	3-6 Month Contract	0.2991	0.3644*
	6-9 Month Contract	0.4010	0.4345**

Note: Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance at 10%, 5%, and 1%, respectively.

Table II-3. Hedging Effectiveness and Optimal Hedging Ratio for DAP

Hedge Period	The Futures Contract	Hedging Effectiveness	Optimal Hedging Ratio
Week Hedges	Tulsa (125 observations)		
	The Nearby Contract(1-3 Month)	0.1625	0.2041**
	3-6 Month Contract	0.1651	0.2852**
	6-9 Month Contract	0.0646	0.1403
	New Orleans and L.A. (125 observations)		
	The Nearby Contract(1-3 Month)	0.2289	0.2689***
	3-6 Month Contract	0.1446	0.2313*
	6-9 Month Contract	0.0425	0.0932
	Central Florida (125 observations)		
	The Nearby Contract(1-3 Month)	0.1149	0.1221
	3-6 Month Contract	0.1155	0.1703
	6-9 Month Contract	-0.0118	-0.0232
Month Hedges	Tulsa (37, 35, and 33 observations)		
	The Nearby Contract(1-3 Month)	0.2606	0.4298
	3-6 Month Contract	0.2414	0.6000
	6-9 Month Contract	-0.0109	-0.0273
	New Orleans and L.A. (37, 35, and 33 observations)		
	The Nearby Contract(1-3 Month)	0.2896	0.5167*
	3-6 Month Contract	0.3443	0.9114**
	6-9 Month Contract	0.0402	0.1198
	Central Florida (37, 35, and 33 observations)		
	The Nearby Contract(1-3 Month)	0.2961	0.4378*
	3-6 Month Contract	0.3386	0.7609**
	6-9 Month Contract	-0.0339	-0.0903

Note: Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance at 10%, 5%, and 1%, respectively.

Table II-4. Week Hedging Effectiveness and Optimal Hedging Ratio for Heating Oil and Diesel

The Futures Contract	Hedging Effectiveness	Optimal Hedging Ratio
	Heating oil in New York Harbor (156 observations)	
1 Month Contract	0.7598	0.7456^{***}
3 Month Contract	0.8765	0.9389^{***}
6 Month Contract	0.8885	1.0491^{***}
9 Month Contract	0.8777	1.1720^{***}
	Heating oil in U.S. Gulf Coast (156 observations)	
1 Month Contract	0.6756	0.6800^{***}
3 Month Contract	0.7814	0.8541^{***}
6 Month Contract	0.7580	0.9972^{***}
9 Month Contract	0.7585	1.0999^{***}
	Diesel in New York Harbor (156 observations)	
1 Month Contract	0.7352	0.7396^{***}
3 Month Contract	0.8514	0.9326^{***}
6 Month Contract	0.8753	1.1044^{***}
9 Month Contract	0.8626	1.2152^{***}
	Diesel in U.S. Gulf Coast (156 observations)	
1 Month Contract	0.6393	0.7741^{***}
3 Month Contract	0.7698	1.0148^{***}
6 Month Contract	0.7695	1.1685^{***}
9 Month Contract	0.7659	1.2986^{***}
	Diesel in L.A. (156 observations)	
1 Month Contract	0.6174	0.6820^{***}
3 Month Contract	0.7080	0.8515^{***}
6 Month Contract	0.7199	0.9972^{***}
9 Month Contract	0.7204	1.1135^{***}

Note: Triple asterisk(***) denote significance at 1%.

Table II-5. Month Hedging Effectiveness and Optimal Hedging Ratio for Heating Oil and Diesel

The Futures Contract	Hedging Effectiveness	Optimal Hedging Ratio
Heating oil in New York Harbor (39 observations)		
1 Month Contract	0.9271	0.9422***
3 Month Contract	0.9544	1.0158***
6 Month Contract	0.9324	1.0312***
9 Month Contract	0.8817	1.0561***
Heating oil in U.S. Gulf Coast (39 observations)		
1 Month Contract	0.9081	0.9851***
3 Month Contract	0.9449	1.0735***
6 Month Contract	0.9276	1.0951***
9 Month Contract	0.8464	1.0821***
Diesel in New York Harbor (39 observations)		
1 Month Contract	0.8947	1.0410***
3 Month Contract	0.9455	1.1522***
6 Month Contract	0.9232	1.1689***
9 Month Contract	0.8597	1.1789***
Diesel in U.S. Gulf Coast (39 observations)		
1 Month Contract	0.8361	1.0407***
3 Month Contract	0.9037	1.1779***
6 Month Contract	0.9017	1.2214***
9 Month Contract	0.8203	1.2033***
Diesel in L.A. (39 observations)		
1 Month Contract	0.7595	0.9332***
3 Month Contract	0.8267	1.0637***
6 Month Contract	0.8418	1.1255***
9 Month Contract	0.7900	1.1439***

Note: Triple asterisk(***) denote significance at 1%.

Table II-6. Impact of Mechanical Purchase Strategies on Average Fertilizer Price

Purchase Strategies	Urea-Tulsa	Urea-Enid	UAN-Tulsa	UAN-Enid	DAP-Tulsa
25% for pre-purchase	.98	.98	.96	.97	.97
50% for pre-purchase	.97	.97	.95	.96	.97
75% for pre-purchase	.96	.96	.94	.94	.96
100% for pre-purchase	.96	.95	.93	.93	.96
Even weekly	1.00	1.00	1.00	1.00	1.00

Price shown relative to a base strategy of purchasing an even amount each week

Table II-7. Impact of Mechanical Purchase Strategies on Price Variance for Various Warehouse Capacity, Locations and Product Forms

Purchase Strategies	Urea-Tulsa	Urea-Enid	UAN-Tulsa	UAN-Enid	DAP-Tulsa
25% for pre-purchase	1.07	1.03	.82	.85	.70
50% for pre-purchase	1.05	.99	.72	.73	.55
75% for pre-purchase	.85	.83	.68	.69	.47
100% for pre-purchase	.74	.83	.66	.67	.43
Even weekly	1.00	1.00	1.00	1.00	1.00

Price shown relative to a base strategy of purchasing an even amount each week

Table II-8. Difference between the Minimum and Maximum Average Fertilizer Price for Various Warehouse Capacity, Locations and Product Forms

Purchase Strategies	Urea-Tulsa	Urea-Enid	UAN-Tulsa	UAN-Enid	DAP-Tulsa
25% for pre-purchase	.08	.08	.10	.07	.07
50% for pre-purchase	.10	.10	.12	.10	.08
75% for pre-purchase	.12	.12	.14	.12	.09
100% for pre-purchase	.13	.14	.16	.15	.09

Price shown relative to a base strategy of purchasing an even amount each week

Table II-9. Difference between the Minimum and Maximum Fertilizer Price Variance for Various Warehouse Capacity, Locations and Product Forms

Purchase Strategies	Urea-Tulsa	Urea-Enid	UAN-Tulsa	UAN-Enid	DAP-Tulsa
25% for pre-purchase	.29	.42	.25	.21	.40
50% for pre-purchase	.35	.44	.71	.30	.52
75% for pre-purchase	.54	.47	.82	.63	.58
100% for pre-purchase	.55	.54	.83	.65	.64

Price shown relative to a base strategy of purchasing an even amount each week

Table II-10. Optimal Time Periods to Purchase Fertilizer

Strategies	Urea-Tulsa	Urea-Enid	UAN-Tulsa	UAN-Enid	DAP-Tulsa
Minimum Average Price	2nd week of July	1st week in July	2nd week of November	2nd week of November	1st week in November
Maximum Average Price	4th week in March	1st week in April	4th week in April	4th week in April	4th week of March
Minimum Variance	Varying amounts over 50 weeks of the year	Varying amounts over 49 weeks of the year	4th week of November	2nd week in January plus 4th week in November	2nd week in January plus 2nd week of November
Maximum Variance	4th week of October	1st week in April	4th week of April	4th week in April	4th week of March

CHAPTER III

ESTIMATING EFFICIENCY WITH STOCHASTIC FRONTIER COST FUNCTION AND AGGREGATE DATA: A MONTE CARLO STUDY

Introduction

Since Farrell (1957) developed his efficiency index using a deterministic frontier function, efficiency measurements from a stochastic frontier model and data envelopment analysis (DEA) have been consistently developed by researchers in order for better efficiency measurement in a specific case. A stochastic frontier function is modeled by Aigner et al. (1977) who brought about the possibility that deviations from the frontier might arise because of random factors and provided the disturbance term as the sum of symmetric normal and half-normal random variables. Later, firm-specific inefficiency measurement was introduced by Jondrow et al. (1982) based on the expected value of an inefficiency error conditional on the overall errors. DEA was introduced by Charnes et al. (1978) with constant returns to scale (CRS) and developed by Banker et al. (1984) extending variable returns to scale (VRS). The main advantage of using DEA is that since it is a nonparametric method it does not need distributional assumptions on error terms. Also, DEA can handle multiple outputs and inputs. On the other hand, the fundamental merit of using a stochastic frontier function is to measure inefficiency in the presence of statistical noise, but this approach needs to assume error structures.

One concern is possible heteroscedasticity. Caudill and Ford (1993) found the biases in the frontier estimation due to heteroscedasticity of a one-sided error and later Caudill, Ford, and Gropper (1995) found that the rankings of firms by efficiency measures were significantly affected by the correction for heteroscedasticity. These followed Schmidt's suggestion (1986) that a one-sided error can be associated with factors under the control of the firm while the random component can be associated with factors outside the control of the firm. More specifically, he argues that heteroscedasticity affected by size is involved in a one-sided error because firm-level data are used in the frontier function and firms vary widely in size which is a factor under the firm's control. Later, Hadri (1999) suggests heteroscedasticity of both error terms with the same data of Caudill, Ford, and Gropper (1995).

Most common occurrence of heteroscedasticity is, in general, when data are aggregated, which is called "groupwise heteroscedasticity (Greene, 2003)". Dickens (1990) showed that aggregated data caused heteroscedasticity with a group size in the presence of a group specific error component so that suggested that using data weighted by the square root of group size was only appropriate if individual error terms are not correlated within groups. In reality, disaggregated data are not available so that economic research is sometimes done by using aggregated data, e.g., hog slaughter industry introduced by Macdonald et al. (2000) and poultry processing industry by Ollinger et al (2005). They use average cost functions and both conclude that small size firms are much more inefficient than large firms. When the average cost function with aggregated data is estimated, the variation of average cost is decreasing as the group size increases so that it brings about the argument of economies of size especially in the estimation of frontier functions.

This paper estimates a stochastic frontier cost function with heteroscedasticity caused by data aggregation and inefficiency indexes from both heteroscedasticity and homoscedasticity by using a Monte Carlo study. We specify the aggregated model from the disaggregated random effects model and take the log of aggregated model, which is the most common way. To make the equation be simplified, first order Taylor approximation is applied, and then heteroscedasticity on error terms is shown. A Monte Carlo study enables to verify it and compare inefficiency measurement in the presence of heteroscedasticity with that of disregarding heteroscedasticity. With the same data, inefficiency indexes from DEA are measured in order for comparisons with the stochastic frontier function.

Theory

Consider the following disaggregated cost function with random effects:

(III-1) $C_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_j + w_{ij}, \quad i = 1, \dots, n_j, \ j = 1, \dots, J,$

$$u_j \sim iid\, N(0, \sigma_u^2),\ w_{ij} \sim iid\, N(0, \sigma_w^2),\ \operatorname{cov}(u_j, w_{ij}) = 0,$$

where C_{ij} is the cost of the *i*th unit in the *j*th firm, $\mathbf{x}_{ij}$ is a vector of explanatory variables including input prices, $\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated, u_j is the random effect of the *j*th firm, w_{ij} is the unexplained portion of the cost of *i*th unit in the *j*th firm.

There are several examples of above disaggregated cost function which can be applied to, e.g., a firm having packing plants, a farmer having many fields, a school having many teachers and etc. A certain firm's disaggregated cost consists of each packing plant's cost. A certain farmer's disaggregated cost consists of each field's cost. A certain school's disaggregated cost consists of each teacher's cost. The unit means each packing plant, each field, and each teacher.

In a stochastic frontier cost function, the inefficiency is considered as the deviations from the frontier so that a one-sided error term is needed to represent the inefficiency (Aigner, Lovell, and Schmidt, 1977). Thus a stochastic frontier cost function can be defined as

$$\text{(III-2)} \qquad C_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_j + w_{ij} + v_j\,,\ v_j \sim iid\left|N(0,\, \sigma_v^2)\right|, \operatorname{cov}(u_j, v_j) = 0\,,$$

$$\operatorname{cov}(w_{ij}, v_j) = 0\,,$$

where v_j is the inefficiency and a one-sided error with $\mathrm{E}(v) = \sigma_v \sqrt{2/\pi}$ and $\mathrm{Var}(v) = \sigma_v^2(1 - 2/\pi)$. Especially, $\sigma_v \sqrt{2/\pi}$ is known as average inefficiency measurement by Aigner et al. (1977). Since the inefficiency error is incorporated, the term ($\mathbf{x}'_{ij}$) can be interpreted as a minimum cost.

When being summed over all outputs within each firm, a (total) stochastic frontier cost function can be derived as

$$\text{(III-3)} \quad \sum_{i=1}^{n_j} C_{ij} = \sum_{i=1}^{n_j} \mathbf{x}'_{ij}\boldsymbol{\beta} + n_j u_j + \sum_{i=1}^{n_j} w_{ij} + n_j v_j\,.$$

where n_j is the number of units produced by *j*th firm.

The dot is the common notation to denote that the variable has been averaged over the corresponding index. A (total) stochastic frontier cost function (III-3) using the dot notation can be also rewritten as

$$\text{(III-4)} \quad TC_j = n_j \mathbf{x}'_{\cdot j}\boldsymbol{\beta} + n_j(u_j + w_{\cdot j} + v_j)\,,\ j = 1, \ldots, J\,,\ w_{\cdot j} \sim N(0, \frac{\sigma_w^2}{n_j}),$$

where TC_j is the total cost for the *j*th firm, and the dot in subscript indicates that the variable has been averaged over units, $\mathbf{x}_{\cdot j}$ is the averaged vector of explanatory variables over units, $w_{\cdot j}$ is the averaged unexplained error over units.

Here, heteroscedasticity related with units is shown, which is typically called groupwise heteroscedasticity (Greene, 2003). Similar heteroscedasticity has been shown by Dickens (1990) in the presence of firm specific error like the random effect here.

A translog cost function is usually used (Melton and Huffman, 1995) due to several conveniences such as including multiple outputs, calculating elasticities easily, adjusting heteroscedasticity, etc. Taking the natural log of equation (III-4) gives

(III-5) $\ln TC_j = \ln n_j + \ln\left(\mathbf{x}'_{.j}\boldsymbol{\beta} + u_j + w_{.j} + v_j\right)$,

which is the most common double log cost function. Then, since error terms are the only random variables, applying first order Talyor approximation of $\ln\left(\mathbf{x}'_{.j}\boldsymbol{\beta} + u_j + w_{.j} + v_j\right)$ around the mean of the random and unexplained error, and the frontier of inefficiency error such as $u_j = 0$, $w_{.j} = 0$ and $v_{.j} = 0$ will give us the following model:

(III-6) $\ln TC_j \approx \ln n_j + \ln \mathbf{x}'_{.j}\boldsymbol{\beta} + \frac{1}{\mathbf{x}'_{.j}\boldsymbol{\beta}}\left(u_j + w_{.j} + v_j\right)$.[1]

The variance of all error terms is $\left(1/\mathbf{x}'_{j}\boldsymbol{\beta}\right)^2\left(\sigma_u^2 + \sigma_w^2/n_j + \sigma_v^2\right)$, shows that explanatory variables including true parameters affect heteroscedasticity. In terms of size-related heteroscedasticity (Schmidt, 1986), the variation of unexplained error is only affected by output level because it is averaged over outputs.

[1] Second order Taylor approximation will give us the following model:
$\ln TC_j \approx \ln n_j + \ln \mathbf{x}'_{.j}\boldsymbol{\beta} + \frac{1}{\mathbf{x}'_{.j}\boldsymbol{\beta}}\left(u_j + w_{.j} + v_j\right) - \frac{1}{2}\left(\frac{1}{\mathbf{x}'_{.j}\boldsymbol{\beta}}\left(u_j + w_{.j} + v_j\right)\right)^2$. This is not considered here since the primary concern is to investigate heteroscedasticity easily in the stochastic frontier function. Also, first order Taylor approximation for an average cost function can be expressed as $\ln AC_j \approx \ln \mathbf{x}'_{.j}\boldsymbol{\beta} + \frac{1}{\mathbf{x}'_{.j}\boldsymbol{\beta}}\left(u_j + w_{.j} + v_j\right)$, which has the same error structure.

When letting $e = w + v$ with an unexplained error (w) and an inefficiency error (v), the density function by Weinstein(1964) is known as

(III-7) $$f(e) = \frac{2}{\sigma} f^{*}\left(\frac{e}{\sigma}\right) F^{*}\left(\frac{\lambda e}{\sigma}\right), -\infty < e < +\infty,$$

where $\sigma^2 = \sigma_w^2 + \sigma_v^2$, $\lambda = \sigma_v / \sigma_w$, f^* and F^* are the standard normal probability density function and the standard normal cumulative density function, respectively.

Here, λ is an indicator of the relative variability of error terms. As Aigner et al. (1977) aregued, $\lambda \to 0$ means $\sigma_v \to 0$ and/or $\sigma_w \to \infty$, i.e. that inefficiency error is dominated by random error.

There are two measurements for the firm-specific inefficiency given by Jondrow et al. (1982). Both are based on the conditional distribution of inefficiency error (v) given overall error (e). The first measure is given by

(III-8) $$\mathrm{E}(v | e) = \sigma_* \left[\left(\frac{\lambda e}{\sigma}\right) + f^{*}\left(\frac{\lambda e}{\sigma}\right) \Big/ F^{*}\left(\frac{\lambda e}{\sigma}\right)\right],$$

where $\sigma_*^2 = (\sigma_v \sigma_w / \sigma)^2$, others are the same definitions as in equation (III-7).

The second measure which is based on the conditional mode is given by

(III-9) $$\begin{aligned} \mathrm{M}(v | e) &= e\left(\sigma_v^2 / \sigma^2\right) \quad \text{if } e \geq 0 \\ &= 0 \qquad\qquad\quad \text{if } e < 0. \end{aligned}$$

Regarding heteroscedasticity and random effects in equation (III-6), the log-likelihood function applied from the density function in the equation (III-7) can be expressed as

(III-10) $$\sum_{j=1}^{J} \ln(f_j(e_j)) = \sum_{j=1}^{J} \ln \frac{2}{\sigma_j} f^{*}\left(\frac{e_j}{\sigma_j}\right) F^{*}\left(\frac{\lambda_j e_j}{\sigma_j}\right),$$

where $e_j = \frac{1}{\mathbf{x}'_{.j}\boldsymbol{\beta}}\left(u_j + w_{.j} + v_j\right)$, $\sigma_j = \left(\frac{1}{\mathbf{x}'_{.j}\boldsymbol{\beta}}\right)\sqrt{\sigma_u^2 + \frac{\sigma_w^2}{n_j} + \sigma_v^2(1-2/\pi)}$, and $\lambda_j = \sqrt{\frac{\sigma_v^2(1-2/\pi)}{\sigma_u^2 + \sigma_w^2/n_j}}$.

This enables using maximum likelihood estimation (MLE) with heteroscedasticity for the stochastic frontier cost function.

Data Envelopment Analysis (DEA)

Since input-oriented efficiency indexes with CRS and VRS were proposed by Charnes et al.(1978) and Banker et al.(1984), respectively, these two technologies have been widely used so that efficiency measurement with CRS and VRS is discussed here (Fare et al., 1994).

Assuming M different outputs, N different inputs, and J different firms, the input-oriented model with VRS can be defined as

(III-11) $$F_j(\mathbf{x}_j, \mathbf{y}_j) = \min_{\theta_j \boldsymbol{\lambda}} \theta_j \quad \text{s.t. } \mathbf{y}_j \le \mathbf{Y}\boldsymbol{\lambda},\ \mathbf{X}\boldsymbol{\lambda} \le \theta_j \mathbf{x}_j,\ \mathbf{j}'\boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda} \ge 0,\ j = 1, \ldots, J$$

where $F_j(\mathbf{x}_j, \mathbf{y}_j)$ is the Farrell efficiency estimate (or technical efficiency) given a $N \times 1$ input vector ($\mathbf{x}_j$) and a $M \times 1$ output vector ($\mathbf{y}_j$) for the j^{th} firm, $\mathbf{Y}$ is a $M \times J$ matrix for outputs, $\mathbf{X}$ is a $N \times J$ matrix for inputs, θ_j is a shrinking factor, $\boldsymbol{\lambda}$ is a $J \times 1$ vector of weights for firms, and $\mathbf{j}$ is a vector of ones.

Farrell efficiency estimate can be measured by the reciprocal of the input distance function since this is input-oriented model. Also, if there is no restriction of $\mathbf{j}'\boldsymbol{\lambda} = 1$, then this is the case of CRS.

Cost minimization model with VRS can be specified as

(III-12)
$$\min_{\boldsymbol{\lambda},\,\mathbf{x}_j^*} \mathbf{r}_j' \mathbf{x}_j^*$$
$$\text{s.t.}\quad \mathbf{y}_j \le \mathbf{Y}\boldsymbol{\lambda},\quad \mathbf{X}\boldsymbol{\lambda} \le \mathbf{x}_j^*,\quad \mathbf{j}'\boldsymbol{\lambda} = 1,\quad \boldsymbol{\lambda} \ge 0,\quad j = 1,\ldots,J,$$

where $\mathbf{r}_j$ is a $N \times 1$ vector of input prices for the j^{th} firm, $\mathbf{x}_j^*$ is the cost-minimizing $N \times 1$ vector of input quantities for the j^{th} firm, which is calculated by the linear programming given a vector of output quantities for j^{th} firm ($\mathbf{y}_j$) and a vector of input prices for j^{th} firm ($\mathbf{r}_j$), and the others are the same as above.

Then, efficiency measurements for the j^{th} firm can be defined as

(III-13)
$$CE = \frac{\text{minimized cost}}{\text{actual cost}} = \frac{\mathbf{r}_j' \mathbf{x}_j^*}{\mathbf{r}_j' \mathbf{x}_j},\quad AE = \frac{CE}{TE},\quad TE = \theta_j,$$

where *CE* is the cost efficiency, *AE* is the allocative efficiency, and *TE* is the technical efficiency derived from the linear programming problem of the equation (III-11).

Data and procedures

A Monte Carlo study can be used to examine heteroscedasticity of a stochastic frontier cost function on error terms. Based on equation (III-2), our true model is assumed as

(III-14)
$$C_{ij} = r_{ij} + u_j + w_{ij} + v_j .$$

where r_{ij} is the input price of the *i*th output in the *j*th firm, the others are the same as previously defined

Aggregation over all outputs yields the following model:

(III-15)
$$\sum_{i}^{n_j} C_{ij} = TC_j = n_j r_{.j} + n_j (u_j + w_{.j} + v_j).$$

Taking a natural log and first order Taylor series around the mean of random errors (u_j and $w_{.j}$) and the frontier (zero) of inefficiency error (v_j) results in the following model:

(III-16) $\ln TC_j \approx \ln(r_{.j}) + \ln n_j + \frac{1}{r_{.j}}(u_j + w_{.j} + v_j)$.

So, our stochastic frontier cost function of equation (III-16) can be rewritten as

(III-17) $\ln TC_j = \beta'_1 \ln r_{.j} + \beta'_2 \ln n_j + (u'_j + w'_{.j} + v'_j)$,

where heteroscedasticity is incorporated into the variances by assuming

$\mathrm{Var}(u'_j + w'_{.j}) = \left(\delta'_u \sigma_u^2 + \delta'_w \frac{\sigma_w^2}{n_j}\right) \frac{1}{(r_{.j})^2}$ and $\mathrm{Var}(v'_j) = \left(\delta'_v \sigma_v^2 (1 - 2/\pi)\right) \frac{1}{(r_{.j})^2}$. Here, β'_1, β'_2, δ'_u, δ'_w, and δ'_v are unknown parameters to be estimated in the presence of heteroscedasticity.

Input prices are generated as $r_{ij} \sim N(12,4)$. Also, lots of small firms and a few large firms are assumed in the unit by using integers of 5*exp(simulated random numbers from a standard normal distribution)+1;; the mean of unit is around 8.89 with variance around 112. In order to see the changes in relative variability of error terms, three scenarios of variances are considered, which are in case of $[\sigma_u^2, \sigma_w^2, \sigma_v^2(1 - 2/\pi)] = [\ 1,\ 4,\ 1.45]$, $[\ 1,\ 4,\ 5.81]$, and $[\ 1,\ 4,\ 13.08]$. Indicators of the relative variability for these are, on average, $\lambda \approx 1$, $\lambda \approx 2$, and $\lambda \approx 3$, respectively. These scenarios show how much the inefficiency indexes for both cases are changed as the variability of inefficiency increases.

Using NLMIXED in SAS with 100 samples of 100 observations, the stochastic frontier cost function with heteroscedasticity and without heteroscedasticity is estimated. Outcomes are first matched with expected values to see how much the model differs from the true model and then outcomes with and without heteroscedasticity are compared.

Since one output and one input are assumed, cost inefficiency is the same as the technical inefficiency from DEA. Inefficiency measurement of DEA using data envelopment analysis program (DEAP) is also calculated and compared with those from the stochastic frontier cost function. Here, constant return to scale (CRS) and variable return to scale (VRS) are applied.

Results

Table III-1 shows mean values of estimated parameters for the stochastic frontier cost function with 100 samples of 100 observations. We estimated $\beta' s$ for the frontier function and $\delta' s$ for the variance equation. Overall, estimated parameters are close to the expected values while the effects of the biases are small.

In terms of average inefficiency from Table III-2, even though inefficiency indexes with heteroscedasticity are slightly bigger than those without heteroscedasticity, which is the same as the previous findings by Caudill, Ford, and Gropper (1995), the differences of inefficiency index between with and without heteroscedasticity is small because of the translog form of the stochastic frontier cost function.

Both unbiasedness and similar inefficiency indexes result from estimation of the translog form of a stochastic frontier cost function. More specifically, first order Taylor approximation gives us heteroscedasticity on inefficiency error term, but the variance of inefficiency error is only affected by inverse of squared explanatory variables including true parameters and if the variation of that value is low enough to be neglected then heteroscedasticity on inefficiency error might be ignored. In our Monte Carlo study, the inverse of squared input price only affects heteroscedasticity and the variation is too small since this is an averaged price over unit. Also, the value itself is too small. Regarding average cost function estimation, these results are expected to be same because it has a same error structure.

Table III-3 and Table III-4 show correlations and rank correlations between true firm specific inefficiency indexes based on conditional mean in equation (8) and firm specific inefficiency indexes from each method. In case of DEA, even though CRS has a higher rank correlation than VAR, this has lower correlations compared to a stochastic frontier function.

Figures III-1 shows the relationship between firm specific inefficiency indexes and output in the 1st sample in case 3. Heteroskedasticity has little impact on firm specific inefficiency index, which is consistent with previous results. However, DEA with VRS tends to have inefficient small firms relative to efficient large firms.

Conclusions

The variations of the stochastic frontier (total) cost function with aggregated data increase as output increases. When the translog form of the stochastic frontier cost function is estimated, all explanatory variables including the true parameters can inversely affect the error terms with first order Taylor approximation. Also, since output affects only the unexplained error term and not the inefficiency error term, if the variations of explanatory variables including true parameters are small, then heteroscedasticity on the inefficiency error might be negligible. In terms of methods, stochastic frontier functions hold up rather well in the presence of data aggregation, but DEA falls apart.

Table III-1. Mean Values from 100 Monte Carlo Trials with First Order Taylor Approximation

Parameters	Case 1			Case 2		
	Expected Values	MLE w/ Hetero	MLE w/ Homo	Expected Values	MLE w/ Hetero	MLE w/ Homo
β_1'	1	1.0207^{***} (0.0220)	1.0235^{***} (0.0222)	1	1.0330^{***} (0.0301)	1.0332^{***} (0.0312)
β_2'	1	1.0020^{***} (0.0152)	1.0032^{***} (0.0160)	1	1.0013^{***} (0.0200)	1.0040^{***} (0.0203)
δ_u'	1	1.4232^{*} (0.5697)		1	2.2525 (1.1257)	
δ_w'	4	4.0161 (2.4111)		4	4.5953 (3.0724)	
δ_v'	1.45	1.3759 (1.3034)		5.81	5.2340 (3.1574)	
$\mathrm{Var}(u'+w')$	0.01	0.0160 (0.0043)	0.0166 (0.0048)	0.01	0.0210 (0.0085)	0.0215 (0.0092)
$\mathrm{Var}(v')$	0.01	0.0098 (0.0093)	0.0083 (0.0086)	0.04	0.0372 (0.0226)	0.0355 (0.0227)

Note: Simulated standard errors are reported in parentheses. Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance on average at 10%, 5%, and 1%, respectively.

1) Case 1 is the case of $[\sigma_u^2, \sigma_w^2, \sigma_v^2(1-2/\pi)] = [1,\ 4,\ 1.45\ ]$.

2) Case 2 is the case of $[\sigma_u^2, \sigma_w^2, \sigma_v^2(1-2/\pi)] = [1,\ 4,\ 5.81]$.

Table III 1. Mean Values from 100 Monte Carlo Trials with First Order Taylor Approximation (Cont.)

Parameters	Case 3		
	Expected Values	MLE w/ Hetero	MLE w/ Homo
β_1'	1	1.0415*** (0.0366)	1.0418*** (0.0419)
β_2'	1	1.0008*** (0.0244)	1.0052*** (0.0256)
δ_u'	1	2.9857 (1.7572)	
δ_w'	4	5.5293 (3.8383)	
δ_v'	13.08	10.7124 (5.2404)	
$\mathrm{Var}(u' + w')$	0.01	0.0255 (0.0131)	0.0268 (0.0146)
$\mathrm{Var}(v')$	0.09	0.0760 (0.0374)	0.0746 (0.0415)

Note: Simulated standard errors are reported in parentheses. Asterisk(*), double asterisk(**), and triple asterisk(***) denote significance on average at 10%, 5%, and 1%, respectively.

1) Case 3 is the case of $[\sigma_u^2, \sigma_w^2, \sigma_v^2(1-2/\pi)] = [1,\ 4,\ 13.08]$.

Table III-2. Mean of Average Inefficiency from 100 Monte Carlo Trials

Methods	Case 1	Case 2	Case 3
True	0.1330	0.2656	0.3990
MLE w/ Hetero	0.1165 (0.0590)	0.2402 (0.0835)	0.3500 (0.0988)
MLE w/ Homo	0.1058 (0.0592)	0.2343 (0.0876)	0.3456 (0.1068)

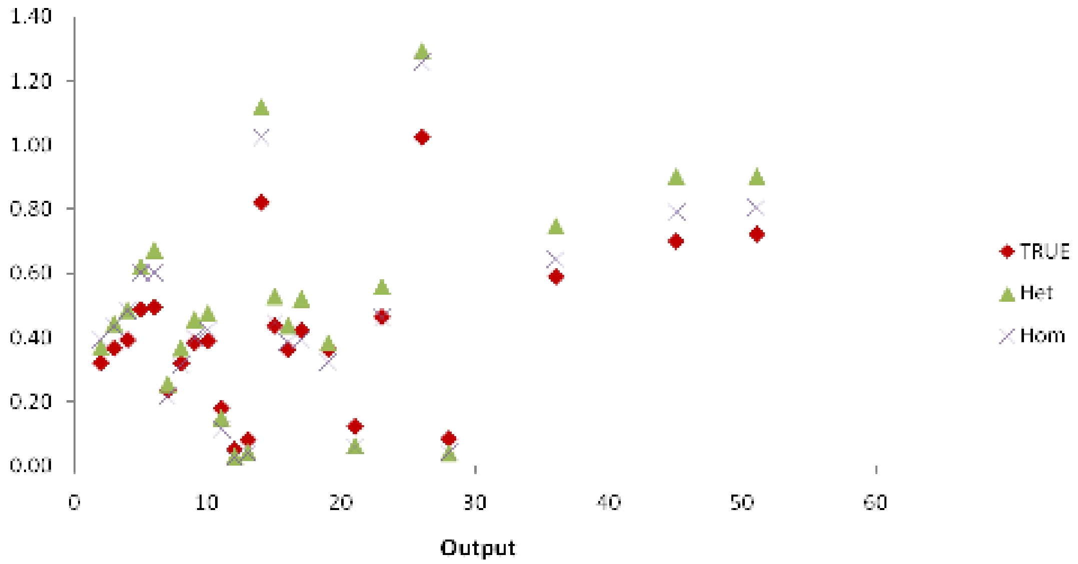
Note: Simulated standard errors are reported in parentheses.

Table III-3. Correlations with True Inefficiency Based on Conditional Mean Using Firm Specific Inefficiency from 100 Monte Carlo Trials

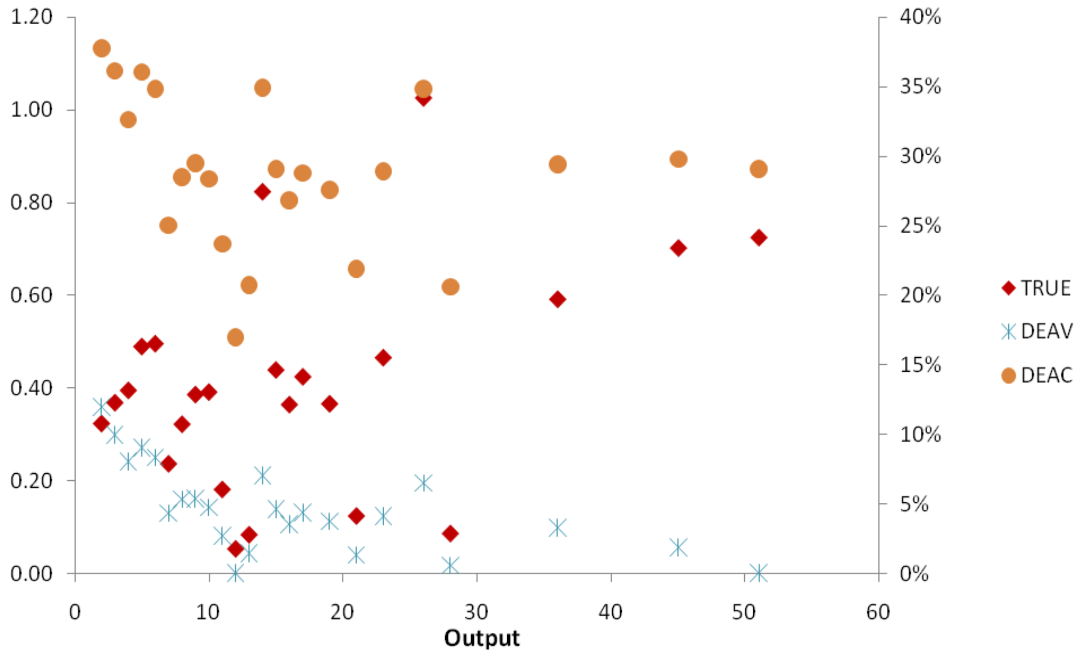
Methods	Case 1	Case 2	Case 3
MLE w/ Hetero	0.9155	0.9259	0.9331
MLE w/ Homo	0.8959	0.9217	0.9265
DEA-CRS	0.3767	0.4839	0.5614
DEA-VRS	0.4044	0.5202	0.5899

Table III-4. Rank Correlations with True Inefficiency Based on Conditional Mean Using Firm Specific Inefficiency from 100 Monte Carlo Trials

Methods	Case 1	Case 2	Case 3
MLE w/ Hetero	0.9784	0.9876	0.9891
MLE w/ Homo	0.9717	0.9880	0.9914
DEA-CRS	0.8518	0.8371	0.8300
DEA-VRS	0.5720	0.6537	0.6985



(a) Inefficiency Index from Stochastic Frontier Cost Function



(b) Inefficiency Index from DEA of Constant Return to Scale and Variable Return to Scale

Figure III-1. Firm specific inefficiency index over averaged output in the 1st sample in case 3

APPENDIX

A. The closed skew-normal (CSN) distribution

If the inefficiency error term (v) in equation (2) is assumed to be varying by not only the j^{th} firm but also i^{th} unit, then the average of sum of CSN distribution is needed. This is the proof and would need to be verified by simulation later.

From Appendix Theorem 1 in Domininguez-Molina et al. (2003, p.9), we have $e = w + v$, $w \sim N(0,\, \sigma_w^2)$, and $v \sim N^0(0,\, \sigma_v^2)$. The closed skew-normal distribution (e) can be expressed as

$$\text{(A-1)} \quad e \sim CSN_{1,1}\left(0,\, \sigma_w^2 + \sigma_v^2,\, \frac{\sigma_v^2}{\sigma_w^2 + \sigma_v^2}, 0,\, \frac{\sigma_w^2 \sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right).$$

Then, the p.d.f. of e is defined as

$$\text{(A-2)} \quad f(e) = \Phi^{-1}\left(0;\, 0,\, \sigma_v^2\right)\phi\left(e;\, 0,\, \sigma_w^2 + \sigma_v^2\right)\Phi\left(\left(\frac{\sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right)\cdot e;\; 0,\, \frac{\sigma_w^2 \sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right).$$

Using Theorem 4 in Gonzalez-Farias et al. (2004, p. 530) and examples of sums of skew-normal random vectors in Genton (2004, p. 37-38) written by Azzalini at al. (1996), the sum of e can be written as

$$\text{(A-3)} \quad \sum_{i=1}^{n} e_i \sim CSN_{1,n}\left(0, n(\sigma_w^2 + \sigma_v^2), \frac{1}{n} D^*, \nu^*, \Delta^*\right),$$

where $D^*_{n\times 1} = \left(\frac{\sigma_v^2}{\sigma_w^2 + \sigma_v^2}, \cdots, \frac{\sigma_v^2}{\sigma_w^2 + \sigma_v^2}\right)^T$, $\nu^*_{n\times 1} = (0, \cdots, 0)^T$, and

$$
\Delta^{*} = \frac{1}{n}\begin{bmatrix}
\frac{\sigma_v^2(n\sigma_w^2+(n-1)\sigma_v^2)}{\sigma_w^2+\sigma_v^2} & -\frac{\sigma_v^2\sigma_v^2}{\sigma_w^2+\sigma_v^2} & \cdots & -\frac{\sigma_v^2\sigma_v^2}{\sigma_w^2+\sigma_v^2} \\
-\frac{\sigma_v^2\sigma_v^2}{\sigma_w^2+\sigma_v^2} & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & -\frac{\sigma_v^2\sigma_v^2}{\sigma_w^2+\sigma_v^2} \\
-\frac{\sigma_v^2\sigma_v^2}{\sigma_w^2+\sigma_v^2} & \cdots & -\frac{\sigma_v^2\sigma_v^2}{\sigma_w^2+\sigma_v^2} & \frac{\sigma_v^2(n\sigma_w^2+(n-1)\sigma_v^2)}{\sigma_w^2+\sigma_v^2}
\end{bmatrix}.
$$

Then, the average of the sum of e and its p.d.f. can be defined as

(A-4) $\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} \sim CSN_{1,n}\left(0, \frac{1}{n}(\sigma_w^2+\sigma_v^2), D^{*}, \nu^{*}, \Delta^{*}\right)$, and

(A-5) $f(\bar{e}) = \Phi_n^{-1}\left(0; \nu^{*}, \Delta^{*} + D^{*}\sigma_w^2 D^{*}\right)\phi\left(\bar{e}; 0, \frac{\sigma_w^2+\sigma_v^2}{n}\right)\Phi_n\left(D^{*}\bar{e}; \nu^{*}, \Delta^{*}\right).$

For example, if n=2 then the mean of the sum of e and its p.d.f. will be written as

(A-6) $\bar{e}_2 = \frac{\sum_{i=1}^{2} e_i}{2} \sim CSN_{1,2}\left(0, \frac{1}{2}(\sigma_w^2+\sigma_v^2), D_2^{*}, \nu_2^{*}, \Delta_2^{*}\right)$, and

(A-7) $f(\bar{e}_2) = \Phi_2^{-1}\left(\begin{bmatrix}0\\0\end{bmatrix}; \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_v^2 & 0\\ 0 & \sigma_v^2\end{bmatrix}\right)\phi\left(\bar{e}_2; 0, \frac{\sigma_w^2+\sigma_v^2}{2}\right)\Phi_2\left(D_2^{*}\bar{e}; \nu_2^{*}, \Delta_2^{*}\right).$

REFERENCES

Aigner, D., C.A.K. Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Models." *Journal of Econometrics* 6: 21-37.

Atlason, J., M. A. Epelman, and S.G. Henderson. 2008. "Optimizing Call Staffing Using Simulation and Analytic Center Cutting-Plane Methods." *Management Science* 54(2): 295-309.

Banker, R.D., A. Charnes, and W.W. Cooper. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30: 1078-1092.

Bollman, K., S. Thompson, and P. Garcia. 1996. "What Killed the Diammonium Phosphate Futures Contract?" *Review of Agricultural Economics*. 25(2): 483-505.

Caudill, S.B., and J.M. Ford. 1993. "Biases in Frontier Estimation Due to Heteroscedasticity." *Economics Letters* 41: 17-20.

Caudill, S.B., J.M. Ford, and D.M. Gropper. 1995. "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity." *Jounal of Business & Economic Statistics* 13(1): 105-111.

Charnes, A., W.W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2: 429-444.

Cherubini, U., E. Luciano, and W. Vecchiato. 2004. *Copula Methods in Finance*. West Sussex: John Wiley & Sons Ltd.

Dickens, W.T. 1990. "Error Components in Grouped Data: Is It Ever Worth Weighting." *The Review of Economics and Statistics* 72(2): 328-333.

Dhuyvetter, K. C., M. Albright, and J. L. Purcell. 2001. "Forecasting and Hedging Crop Input Prices" Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, IL 2001 [http://www.farmdoc.uiuc.edu/nccc134].

Dominguez-Molina, J. A., G. Gonzalez-Farias, and R. Ramos-Quiroga. 2003. "Skew-Normality in Stochastic Frontier Analysis." *Comunicacion Tecnica* I-03-18: 1-13.

Ederington, L. H. 1979. "The Hedging Performance of the New Futures Markets." *Journal of Finance* 34: 157-170.

Fare, R., S. Grosskopf, and C. A. K. Lovell. 1994. *Production Frontiers.* Cambridge University Press.

Farrell, M.J. 1957. "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society. Series A(General)* 120: 253-90.

Genest, C., K. Ghoudi, and L. -P. Rives. 1995. "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions." *Biometrika* 82: 543-552.

Genton, M. G. 2004. *Skew-Elliptical Distributions and Their Applications: A Journey beyond Normality*. Florida: Chapman & Hall/CRC.

Gonzalez-Farias, G., A. Dominguez-Molina, and A. K. Gupta. 2004. "Additive Properties of Skew Normal Random Vectors." *Journal of Statistical Planning and Inference* 126: 512-534.

Green, L. V., P. J. Kolesar, and J. Soares. 2001. "Improving the SIPP Approach for Staffing Service Systems That Have Cyclic Demands." *Operations Research.* 49(4): 549-564.

Greene, W.H. 1980. "On the Estimation of a Flexible Frontier Production Model." *Journal of Econometrics* 13: 101-115.

______. 2003. *Econometric Analysis*, 5th ed. New Jersey: Pearson Education, Inc.

Hadri, K. 1999. "Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function." *Journal of Business & Economic Statistics* 17(3): 359-363

Hann, L. D., and A. Ferreira. 2006. *Extreme Value Theory: An Introduction*. New York: Springer Science+Business Media LLC.

Joe, H., and J.J. Xu. 1996. *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Dept. of Statistics University of British Columbia, Technical Report 166.

Johnson, L. L. 1960. "The Theory of Hedging and Speculation in Commodity Futures." *The Review of Economic Studies* 27(3): 139-151.

Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt. 1982. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics* 19: 233-238.

Laws, Forest "Roller Coaster' Fertilizer Prices Creating Anxiety, *Southwest Farm Press*, Dec 29, 2008 http://southwestfarmpress.com/news/fertilizer-prices-1229/

Liu, W., and J. Cela. 2008. "Count Data Models in SAS." Paper presented at SAS Global Forum 2008, San Antonio TX, 16-19 March.

MacDonald, J.M., and E.O. Michael. 2000. "Scale Economies and Consolidation in Hog Slaughter." *American Journal of Agricultural Economics* 82(2): 334-346.

McNeil A. J. 1997. "Estimating the Tails of Loss Severity Distributions Using Extreme Value Thoery." *ASTIN Bulletin International Actuarial Association* 12(1): 117-137.

Melton, B. E., and W. E. Huffman. 1995. "Beef and Pork Packing Costs and Input Demands: Effect of Unionization and Technology." *American Journal of Agricultural Economics* 77(3): 471-485.

Oklahoma State University, Crop Enterprise Budgets, 2008. http://agecon.okstate.edu/budgets/

Ollinger, M., J.M. MacDonald, and M. Madison. 2005. "Technological Change and Economies of Scale in U.S. Poultry Processing." *American Journal of Agricultural Economics* 87(1): 116-129.

Richardson, J. W., S. L. Klose, and A. W. Gray. 2000. "An Applied Procedure for Estimating and Simulating Multivariate Empirical (MVE) Probability Distributions in Farm-Level Risk Assessment and Policy Analysis." *Journal of Agricultural and Applied Economics* 32(2): 299-315.

Schmidt, P. 1986. "Frontier Production Functions." *Econometric Reviews* 4: 289-328.

Stein, J. J. 1961. "The Simultaneous Determination of Spot and Futures Prices." *American Economic Review* 51: 1012-1025.

Taylor, C. R. 1990. "Two Practical Procedures for Estimating Multivariate Nonnormal Probability Density Functions." *American Journal of Agricultural Economics* 72: 210-217.

USDA National Agricultural Statistics Service "Average U.S. farm prices of selected fertilizers 1960-2008" December, 2008 http://www.ers.usda.gov/Data/FertilizerUse/

Weinstein, M.A. 1964. “Query 2: The Sum of Values from a Normal and a Truncated Normal Distribution.” *Technometrics* 6(1): 104-105.

VITA

TAEYOON KIM

Candidate for the Degree of

Doctor of Philosophy

Dissertation: THREE ESSAYS ON APPLIED ECONOMICS: RURAL ELECTRIC COOPERATIVE CALL CENTER DEMAND, FERTILIZER PRICE RISK, AND ESTIMATING EFFICIENCY WITH DATA AGGREGATION

Major Field: Agricultural Economics

Biographical:

Personal Data: Born in Jinju, Republic of Korea, on October 1, 1975, the son of Jaekwon Kim and Sunja Jung, married with Younghwa Kang on November 27, 2004, and one daughter, Jacqueline Hannah Kim, on March 14, 2007, and one son, Sean Youngwook Kim, on December 30, 2008.

Education: Received Bachelor of Science degree and Master of Science degree in Agricultural Economics from Seoul National University, Republic of Korea in August 2001 and February 2004, respectively. Completed the requirements for the Doctor of Philosophy degree with a major in Agricultural Economics at Oklahoma State University in July, 2009 of Graduation.

Name: TAEYOON KIM Date of Degree: July, 2009

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THREE ESSAYS ON APPLIED ECONOMICS: RURAL ELECTRIC COOPERATIVE CALL CENTER DEMAND, FERTILIZER PRICE RISK, AND ESTIMATING EFFICIENCY WITH DATA AGGREGATION

Pages in Study: 77 Candidate for the Degree of Doctor of Philosophy

Major Field: Agricultural Economics

Scope and Method of Study: This study consists of three sections. The purpose of first section is to forecast peak call volume to allow a centralized after-hours call center for rural electric cooperatives to estimate staffing levels. A Gaussian copula is used to capture the dependence among nonnormal distributions. The purpose of second section is to examine the effectiveness of systematic cash purchase strategies in reducing fertilizer price risk for Oklahoma fertilizer dealers. The historical effectiveness of hedging with the fertilizer future market contracts (which have been discontinued) is also analyzed to provide a benchmark for comparison. The purpose of last section is to determine the effects of data aggregation on estimation of a stochastic frontier cost function using a Monte Carlo study.

Findings and Conclusions: For the first section, ignoring the dependence that the copula includes, would have resulted in an underestimation of peak values. The centralized call center resulted in cost savings of approximately 75% relative to individual centers at each cooperative. Adding cooperatives to the centralized call center is projected to further decrease costs per member. The magnitude of additional cost savings depends on the regional location of the new call center member. For the second section, cash purchase strategies were shown to be slightly effective in reducing average price and moderately effective in reducing risk. The reduction in price variance through cash purchase strategies was comparable to the historical effectiveness of hedging. For the last section, when the translog form of a stochastic frontier cost function with aggregated data is estimated, the variations of total cost decrease as output increases. If the variations of explanatory variables are small, then heteroscedasticity on the inefficiency error might be negligible. Stochastic frontier functions hold up rather well in the presence of data aggregation, but efficiency measurement from DEA diverges from true efficiency measurement.

ADVISER'S APPROVAL: Dr. Philip Kenkel